

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 09:56:27 ; Search time 4995 Seconds
(without alignments)
11483.810 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaaaaagctataaccgc.....catggtttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1971	100.0	1971	6	AX006271	AX006271 Sequence
2	1971	100.0	217420	1	BSUB0017	Z99120 Bacillus su
3	807.8	41.0	1958	6	AX431809	AX431809 Sequence
4	238.2	12.1	265118	1	CNSPAX06	AJ248288 Pyrococcus
5	238.2	12.1	265118	6	AX041922	AX041922 Sequence
6	227.2	11.5	1896	6	AX116209	AX116209 Sequence
7	227.2	11.5	1896	6	E16634	E16634 Pyrococcus
8	227.2	11.5	233000	1	AP000003	AP000003 Pyrococcus
9	218.2	11.1	12882	1	AE013176	AE013176 Thermoana
10	192.2	9.8	477	6	AX436451	AX436451 Sequence
11	191.2	9.7	10256	1	AE001879	AE001879 Deinococc
12	161.8	8.2	251700	1	AP000062	AP000062 Aeropyrum
13	159.6	8.1	9531	1	AE010617	AE010617 Fusobacte
14	158.2	8.0	11233	1	AE009798	AJ248287 Pyrococcus
15	158	8.0	304290	1	CNSPAX05	AX041921 Sequence
16	158	8.0	349980	6	AX041921	AX041921 Pyrococcus
17	150.6	7.6	257000	1	AP000002	AP000002 Pyrococcus
18	149.8	7.6	1869	6	AR086952	AR086952 Sequence
19	149.8	7.6	1869	6	AR095647	AR095647 Sequence
20	149.8	7.6	1869	6	AR136115	AR136115 Sequence
21	143.4	7.3	268050	1	AP000983	AP000983 Sulfolobu
22	143	7.3	12358	1	AE006863	AE006863 Sulfolobu
23	132.6	6.7	182440	1	AP000994	AP000994 Thermopla
24	130.4	6.6	9825	1	AE005114	AE005114 Halobacte
25	130.4	6.6	247695	1	AP000064	AP000064 Aeropyrum
26	128.4	6.5	13354	1	AE010156	AE010156 Pyrococcus
27	128	6.5	11370	1	AE013111	AE013111 Thermoana
28	82.4	4.2	585	6	AX436423	AX436423 Sequence
29	75.2	3.8	139203	2	AC068311	AC068311 Homo sapi
30	73	3.7	12098	1	AE012209	AE012209 Xanthomon
31	65	3.3	1998	3	AF059278	AF059278 Entamoeba
32	64.8	3.3	2400	1	BSTHRC	AF059279 Rickettsi
33	64.2	3.3	12621	1	AE008621	AE008621 Rickettsi
34	62.8	3.2	1992	3	AF059279	AF059279 Entamoeba
35	59.2	3.0	2642	1	AF508804	AF508804 Xanthomon
36	57	2.9	11388	1	AE011750	AE011750 Xanthomon
37	54.4	2.8	7218	6	I66494	I66494 Sequence 14
38	48.4	2.5	11505	1	AE012120	AE012120 Xanthomon
39	46.4	2.4	1797	6	AX431923	AX431923 Sequence
40	46.4	2.4	12345	1	AE006757	AE006757 Sulfolobu
41	46.2	2.3	316900	1	TACID3	AL45065 Thermopla
42	46	2.3	2325	8	AF082514	AF082514 Trichophy
43	45.4	2.3	13928	1	AF004038	AF004038 Xylella f
44	45.2	2.3	782	11	PM11H10G	AL684260 Penicilli
45	45.2	2.3	10393	1	AE011651	AE011651 Xanthomon

ALIGNMENTS

RESULT 1
AX006271
LOCUS AX006271 1971 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 1 from Patent WO9903984.
ACCESSION AX006271
VERSION AX006271.1 GI:9929124
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 1971)
AUTHORS Estell,D.A.
TITLE Proteases from gram-positive organisms
JOURNAL Patent: WO 9903984-A 1 28-JAN-1999;
GENENCO INTERNATIONAL B V (NL); ESTELL DAVID A (US)

FEATURES

Location/Qualifiers

source

1. .1971
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 558 a 431 c 524 g 458 t
ORIGIN

Query Match 100.0%; Score 1971; DB 6; Length 1971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAGCTGATACACCGACGACATCACAGCGATTGTCTCTGTGACCGATCCTCAA 60
DB 1 ATGAAAAGCTGATACACCGACGACATCACAGCGATTGTCTCTGTGACCGATCCTCAA 60

QY 61 TAGGCCCCAGACGGTACCGTCCCGCATATGTAATCAAAATCAAGTAAATCAAGAGAAGAT 120
DB 61 TAGGCCCCAGACGGTACCGTCCCGCATATGTAATCAAAATCAAGTAAATCAAGAGAAGAT 120

QY 121 TCGTATACATCAAAATATATGATCTATGAACGAAACGGGAGGATCTGCTCTTGGACA 180
DB 121 TCGTATACATCAAAATATATGATCTATGAACGAAACGGGAGGATCTGCTCTTGGACA 180

QY 181 CATGAGAAAACGGAACGACCCAGATGGTCTCCGGAGCGGCGACGCTTCGCTTT 240
DB 181 CATGAGAAAACGGAACGACCCAGATGGTCTCCGGAGCGGCGACGCTTCGCTTT 240

QY 241 ATTTCTGATCGAAGAGCGGATCGGCGACAGCTTTATATCATGACACCTGAAGCGGAGAA 300
DB 241 ATTTCTGATCGAAGAGCGGATCGGCGACAGCTTTATATCATGACACCTGAAGCGGAGAA 300

QY 301 GCAAGAAACCTGACTGATATCCATATGCGGTGTCGAGCGGCTATGTCCTCCGACGGT 360
DB 301 GCAAGAAACCTGACTGATATCCATATGCGGTGTCGAGCGGCTATGTCCTCCGACGGT 360

QY 361 GAATCGATTCTGGTCACTATCAGTTGGGAGGAGGAGAAACATGATGACCGGAGAAAA 420
DB 361 GAATCGATTCTGGTCACTATCAGTTGGGAGGAGGAGAAACATGATGACCGGAGAAAA 420

QY 421 ACAGAGGAGACGCTATGAACCTGTGAAGTCAAGGCCCTCTCTCAAAACGGAGCGG 480
DB 421 ACAGAGGAGACGCTATGAACCTGTGAAGTCAAGGCCCTCTCTCAAAACGGAGCGG 480

QY 481 AAAGGCTGACGAGAGTGGCTATGCCAGCTTGTGTCGACGCTAAAGTCCGGTACG 540
DB 481 AAAGGCTGACGAGAGTGGCTATGCCAGCTTGTGTCGACGCTAAAGTCCGGTACG 540

QY 541 ATGAAAGCTGACAACTGATCAAGCTGATCGGTGATCCTGCTTTTCTCTGACGGC 600
DB 541 ATGAAAGCTGACAACTGATCAAGCTGATCGGTGATCCTGCTTTTCTCTGACGGC 600

QY 601 AAATGGCTGTTTCTCAGCTAATTAACCTGAACAGATGATGCCAGCAAGCGCATGAT 660
DB 601 AAATGGCTGTTTCTCAGCTAATTAACCTGAACAGATGATGCCAGCAAGCGCATGAT 660

QY 661 GTTTACATATGCTGAGTCTGAGATCTTAAGCAGGTTACACCTCATCTCGGGCTCA 720
DB 661 GTTTACATATGCTGAGTCTGAGATCTTAAGCAGGTTACACCTCATCTCGGGCTCA 720

QY 721 TTCGGATCAAGCTCATTTTACACAGCGGAGGATCTGCTTTGCTTGGAAATGAAAG 780
DB 721 TTCGGATCAAGCTCATTTTACACAGCGGAGGATCTGCTTTGCTTGGAAATGAAAG 780

QY 781 GAATATAAGATGCTACGCTCTCAAGGGTGGCTCTATGATATCAACAGCGCCGCTC 840
DB 781 GAATATAAGATGCTACGCTCTCAAGGGTGGCTCTATGATATCAACAGCGCCGCTC 840

QY 841 ACATGCTTACTGAGATGCTGACGCTTCAATAGCGGATGCGGTGATGAGATTCATTG 900
DB 841 ACATGCTTACTGAGATGCTGACGCTTCAATAGCGGATGCGGTGATGAGATTCATTG 900

QY 901 ATCGGTGGTCTGAACAGCGCCGATTTGGACAAAGGACAGCGGTTTATCTCATC 960
DB 901 ATCGGTGGTCTGAACAGCGCCGATTTGGACAAAGGACAGCGGTTTATCTCATC 960

RESULT 2
BSUB0017/c

901 ATCGGTGGTCTGAACAGCGCCGATTTGGACAAAGGACAGCCAAAGGTTTATGTCATC 960
961 GGCACAGATCAAGCGAGTACGGCATCTATATATTTTCGATTTGAAGCCCTTGTGATCCG 1020
961 GGCACAGATCAAGCGAGTACGGCATCTATATATTTTCGATTTGAAGCCCTTGTGATCCG 1020

1021 ATTCGTCTGGAAAAGAGTACATCAATAGCTTTCTCTTTCACCTGATGACAGCAGCTTT 1080
1021 ATTCGTCTGGAAAAGAGTACATCAATAGCTTTCTCTTTCACCTGATGACAGCAGCTTT 1080

1081 ATTCGACGTGTGACAAAGCGGACAGACCGAGTGAAGCTTTTACAGTATCCCGCTTGACAG 1140
1081 ATTCGACGTGTGACAAAGCGGACAGACCGAGTGAAGCTTTTACAGTATCCCGCTTGACAG 1140

1141 GAAGAGAAAACAGTCTGCTGCGGAATGACAAAGTTTGTTCAGGAGCATACGATATCAATA 1200
1141 GAAGAGAAAACAGTCTGCTGCGGAATGACAAAGTTTGTTCAGGAGCATACGATATCAATA 1200

1201 CCTGAAGAGATTCAATATGCTACAGAAAGCGGCTGATGTTGAACGGCTGGCTGATGAGG 1260
1201 CCTGAAGAGATTCAATATGCTACAGAAAGCGGCTGATGTTGAACGGCTGGCTGATGAGG 1260

1261 CCTGCACAAATGAAAGGTGAGACAAATATCCACTTATTTTAAACATACACGCGGCTCG 1320
1261 CCTGCACAAATGAAAGGTGAGACAAATATCCACTTATTTTAAACATACACGCGGCTCG 1320

1321 CATATGATGTACGGACATACATATTTTCATGAGTTTTCAGGTGCTGCGGCGGAAAGGATAC 1380
1321 CATATGATGTACGGACATACATATTTTCATGAGTTTTCAGGTGCTGCGGCGGAAAGGATAC 1380

1381 CGGTCGCTTATATCAATCCGAGAGAAAGCCAGGCTACGGGACAGAAATTTGTGATGCGG 1440
1381 CGGTCGCTTATATCAATCCGAGAGAAAGCCAGGCTACGGGACAGAAATTTGTGATGCGG 1440

1441 GTCAGAGAGATTATGGGGAAAGGATTATGACGATGTGATGACGCTGTGGATGAGGCT 1500
1441 GTCAGAGAGATTATGGGGAAAGGATTATGACGATGTGATGACGCTGTGGATGAGGCT 1500

1501 ATCAAGAGAGATCCGATATTTGATCTTAAGCGGCTCGGTGTACGCGGCGGAGCTACCGA 1560
1501 ATCAAGAGAGATCCGATATTTGATCTTAAGCGGCTCGGTGTACGCGGCGGAGCTACCGA 1560

1561 GGTTTTATGACCAACTGGATCGTGGGACAGCAACCGCTTTAAAGCTGCGGTTACCCAG 1620
1561 GGTTTTATGACCAACTGGATCGTGGGACAGCAACCGCTTTAAAGCTGCGGTTACCCAG 1620

1621 CGCTCGATATCAAAATTTGGATCAGCTTTTACGCGGCTGATGATCGGCTATTTCTTTACA 1680
1621 CGCTCGATATCAAAATTTGGATCAGCTTTTACGCGGCTGATGATCGGCTATTTCTTTACA 1680

1681 GACTGGCAGCTTGAGCATGACATGTTTGGAGACACAGAAAGCTCTGGACCGGCTCCT 1740
1681 GACTGGCAGCTTGAGCATGACATGTTTGGAGACACAGAAAGCTCTGGACCGGCTCCT 1740

1741 TTAATATACCGCAACGCTGGAGACACCGCTTTTGTGATGATGCGAGCGGATGAC 1800
1741 TTAATATACCGCAACGCTGGAGACACCGCTTTTGTGATGATGCGAGCGGATGAC 1800

1801 CGATGCCGATCGAGAGCGGAGCAGCTGTTTATCGCTCTGCTGCTGCTGCTGCTGCTGCT 1860
1801 CGATGCCGATCGAGAGCGGAGCAGCTGTTTATCGCTCTGCTGCTGCTGCTGCTGCTGCT 1860

1861 ACCAAGCTTGTCCGTTTCCGAATGATCGCAATTTATACGCGACCGGACCCCAAGA 1920
1861 ACCAAGCTTGTCCGTTTCCGAATGATCGCAATTTATACGCGACCGGACCCCAAGA 1920

1921 CAGCGGATCAAGCGCTGATATATATGATGATGATGATGATGATGATGATGATGATGAT 1971
1921 CAGCGGATCAAGCGCTGATATATATGATGATGATGATGATGATGATGATGATGATGAT 1971

LOCUS	BSUB0017	217420 bp	DNA	linear	BCT 26-NOV-1997
DEFINITION	Bacillus subtilis complete genome (section 17 of 21): from 3197001 to 3414420.				
ACCESSION	Z99120 AL009126				
VERSION	Z99120.1 GI:2635613				
KEYWORDS					
SOURCE	Bacillus subtilis.				
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
REFERENCE	1 (bases 1 to 217420)				
AUTHORS	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Boursier, L., Bressan, P., Brott, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruch, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D., Emerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Gollightly, E. J., Grandi, G., Guisepi, G., Guy, B. J., Haga, K., Halech, J., Harwood, C. R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Ito, Y., Jones, J., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynold, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tanakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H. and Danchin, A.				
TITLE	The complete genome sequence of the gram-positive bacterium				
JOURNAL	Bacillus subtilis				
MEDLINE	Nature 390 (6657), 249-256 (1997)				
PUBMED	98044033				
REFERENCE	9384377				
AUTHORS	2 (bases 1 to 217420)				
TITLE	Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.				
JOURNAL	Direct Submission				
FEATURES	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48				
source	Location/Qualifiers				
	1. .217420				
	/organism="Bacillus subtilis"				
	/strain="168"				
	/db_xref="taxon:1423"				
	complement(3. .1277)				
gene	/gene="yulB"				
	complement(3. .1277)				
CDS	/gene="yulB"				
	/function="unknown"				
	/note="similar to L-rhamnose isomerase"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CAB15107.1"				
	/db_xref="GI:2635614"				
	/db_xref="SPTREMBL:O05264"				
	/translation="MTIKANDSAKQAYEKWIDVEEALROLEQVPSIHWCQGDIE GFVNGKLSGSDVGTGNYPGAQTPPELRDLKALSLIPGKRVNLHAIYAETNRE AVERDLKPOHFENWVWAKNLGLDFNPTFLFSEKAADGLTSHDPDPIREFWIRH CIACRRIGEFYFGKELGTCLTNWIPDGYKDIPDRTPRRLKESLDRIFSEISEQ				
	HNLDSIESKFLGSGSEYVVGSHFEYLAVALTNHKLCLLDTGTHFHTPTVTSNKISSML LYTKDLALHVSRRPVRWSDHVVLDDELRETALEIVRNHAKLEKVAIGLDFDASINRY AAWTIGTRNMKIALYALLPGLNGKLQKEGRTERIALMEEFKTYPPFGAIWDSYCE OMGVPVKEAWLYDIKEYBQQVLLKRASSPIV"				
	complement(1303. .1617)				
gene	/gene="yulD"				
	complement(1303. .1617)				
CDS	/gene="yulD"				
	/function="unknown"				
	/note="similar to hypothetical proteins"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CAB15108.1"				
	/db_xref="GI:2635615"				
	/db_xref="SPTREMBL:O05263"				
	/translation="MKRKASIMFVHODKYEEYKQRHDDIWPEMALKAHGAHHSIF LDFTGRLFAYLEIDEDEKRWKMDTEVCYRWKWSMAPLMKNTSDFSPVAILKEVYF				
	complement(1635. .3092)				
gene	/gene="yulC"				
	complement(1635. .3092)				
CDS	/gene="yulC"				
	/function="unknown"				
	/note="similar to rhamnulokinase"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CAB15109.1"				
	/db_xref="GI:2635616"				
	/db_xref="SPTREMBL:O05262"				
	/translation="MIYTAIDVASSGRIMVGMELNEGKLDIOEIRHFANGFSQRDGH LWDIDHLKILQLOKQVKTGTYHCFTVIGTIDTAVYVLLDEKGRLEAISYDRRT DHTDKLEHTLSKAAIYOKTQIOFQFNTIYOLFEDRELLKTKDKIMMIDYLGCL TGKAVETITVSTQLLAVSTGNLDPLELAVSLEQOQFAPLTPGCEGLKRNWEP DYDLPACKWTVATHDTSASAVIAAPGVNDGWAYISSTGWSLIGVENKTPILDLALE NNYTERANNITIRLKNIIIGHMWIOEYVQQLQADISFOQLAEEAKTTPFQOQFINLD KRFNPNMKEIOHQHYCTQKTPRTAGLACCIYSNLAIYIAIAIKELETTIKPPI EGFHIGGARDFNLQNTADMSKAVYAGPIEATATGNLLMQMTAAKEVKDIKEARQ VVRNSFPKTVFTPKDIDRSTIIQSFQQTVLKLSK"				
	complement(3097. .3873)				
gene	/gene="yulB"				
	complement(3097. .3873)				
CDS	/gene="yulB"				
	/function="unknown"				
	/note="similar to transcriptional regulator (Deor family)"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CAB15110.1"				
	/db_xref="GI:2635617"				
	/db_xref="SWISS-PROT:O05261"				
	/translation="MLVAERQOKIIVEIVNMRSSIRVSELSDIFSVEETIRDLKLE KEHLRSRGGAVSTOQKSEIHFSEIREITNVIEKKAHAHAAYKVGSDRIIDAST TAWYMAKILPDIELTVITNSMKAATLSNKENISVISTGILLESKLSFAGPLAERSL ETYHVNKTFLSCKGFDVHNGSDSNEMOALLKRMERSDOTILLADSSKWKGNREFSH TASQDVSRRLTDSGLDPSAKALEDKVKVTVAPLSKRG"				
	complement(3930. .5999)				
gene	/gene="yuxG"				
	complement(3930. .5999)				
CDS	/gene="yuxG"				
	/function="unknown"				
	/note="alternate gene name: yula				
	similar to sorbitol-6-phosphate 2-dehydrogenase"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CAB15111.1"				
	/db_xref="GI:2635618"				
	/db_xref="SWISS-PROT:P40747"				
	/translation="MVKHWDSERAAQLPKGVVEELVYRNLIGSDRTVCNWWGGNTSM KTKEDFRGRETEVNMVKGWSDSLATMKAHNPSGLKDDIRPLIKRDQMPDEMYDYL SHCMZKHPSPSIETLLHAFYPYKHVDTHPDATISICCADQPDYVGNRFVW VYVPGSTLTKMIAFGVANNPAPHALVMEKHLVTVGTSETCYQKTIISIOEAOY INDRINQHEVGGKRYQPLPDKRQKILAGINPVRGAVSEKKMLISVDDHDDVLEF VNSVQAPALSQIGAAACDHLVTRKRPVLYIDNPETQDVHKDLALIKSGVETFTSEIQ AYFTRNQDQDGQIFESAPRVLIPGIMVNTGKSVAMSKVSGALYRAIVAMKGATL				

GGFVSLHENEYSHYEYWPLEYKLTLPAPAEFFSRKVALITGGAGIGSACRRFAAE
GHRVIVADLNTAQKIAKEINDAYKGRMAKMDVTKEEDVQSAFRAALVAGGID
IVVNAAGLATSPEDETSLEKEMNNMNVLTGTGYFLVAREAFKQKHQRGSMVFGS
KNSVYAGKNASAYS SVKALETHLARCIAAEGGEFGRVNSVLPDVALGSAIWSGSR
EERAAAGIPEPDQLEEHYRKTALLVNIYPEDIAEAIAFFASSKAEKTTGCMITVDGG
VPAAFTR

gene complement(6109. .8125)
/gene="tlpB"
terminator complement(6109. .6124)
/gene="tlpB"
CDS complement(6137. .8125)
/gene="tlpB"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="CAB15112.1"
/db_xref="GI:2635619"
/db_xref="SWISS-PROT:P39217"
/translation="MGKFIOWIKOPISIKPLIAFLAVLILPVGLVAYFSYQSANNAL
DRELISSAGNVEELNSTLONKLEDKVKYADYSETVDKILLGNKNTLLKEKFOYT
TLNDVGAIYAASEDKLYKYPDSVPKGFDPDTGRDWDYQVAEKGQAVFSEPYDEA
TGDIVVTISKQLDGSVIALDLNDELVTASKRIKIGKEGFAPITGNKKYIAHPTI
KPGTTGSDWTNQVYSKEGSEFETFEKKEKMAFTTNKLTGKWTAGTTFVSELODAS
PVLNVAIILCVSIVIGGILILYIIRAITPLRLKLVSTSAKISSGDLTEVIDHSKN
EFGOLGSEFNMSASLSRSLVIGIQTSENVASSEELTASAAQTSKATEHITLAEQF
SDGNEAQSEKLETSNNLSQMNNEGTSKVAQASSTITKSSIQSSPAAGSGEKLVEHTVG
QMKTIQDSVOKAEAVKGLTQSODITSLVINGIADOTNLLALNAIEARAGEYG
RGFSVAEEYKRLAVQSADSAKEIEGLIOEIVRELSLSPFQSVNHBKVGLOITDQ
TASFKQIYEMTQISGELONLNTVQLSAGSEVSSAEDISAVAKESSAGIODIA
ASAEQLASMEETSSAETLANMAEELQDIITKKFKIES"
complement(8204. .10224)
/gene="mcpA"
terminator complement(8204. .8228)
/gene="mcpA"
CDS complement(8239. .10224)
/gene="mcpA"
/function="chemotaxis deficiency toward glucose and
alpha-methyl-glucoside"
/alpha_start=1
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="CAB15113.1"
/db_xref="GI:2635620"
/db_xref="SWISS-PROT:P39214"
/translation="MKKILQIKORSITRKLVSFLSILIPVILAFAYQSASSSL
DROMGSALENVOOLNEIINTSICEKENSADYFSEWLTKEKYNKASINAREKFSQYI
SINKDSEIVTSDTKGHETRYPDLPMPSGYNPVERDWTAKAVANGKVITDPYKTAS
TNWVVTIAQTRKDGSGVIALNMTIENLLTKTKVNGTQGYAFIMTKDKKVAHPNE
QSGTELKGDWLDKMSADKGFQITMDGDKKMAFDNKLTKGKIGGTYLDEIHEAA
QPVHLALIVLAAAIIGIIVMTLIIIRSIITPLKOLVSSKRISEGDLTETIDRSKD
ELGELGKSFNMSASLSRSLHAIQDSVDNVAASSEELTASAAQTSKATEHITLAEQF
SNGEKONENIETAAEHYQMNGLTNMAQASEVITDSSVOSTEIASBGLVHQTG
QMNVIDKSVKEAEQVRLGTEKSKDITNLRVINGIADOTNLLALNAIEARAGEYG
RGFSVAEEYKRLAVQSADSAKEIEGLIIEIVKEINTSLGFMFQSVNQBVGGLDITDK

Query Match 100.0%; Score 1971; DB 1; Length 217420;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAAAGCTGATACCGCAGAGGACATCACCGGATTGTCTGTGACCGATCCTCAA 60
Db 114794 ATGAAAAGCTGATACCGCAGAGGACATCACCGGATTGTCTGTGACCGATCCTCAA 114735
QY 61 TAGCCCCAGACGTCACCGTCGCGCATATGTAAATACAAAGTAAATCAAGAGAAAGAT 120
Db 114734 TAGCCCCAGACGTCACCGTCGCGCATATGTAAATACAAAGTAAATCAAGAGAAAGAT 114675
QY 121 TCGTATACATCAATATATGATCTATGAACGAAAAACGGGAGGATCTGTTCCTTGGACA 180
Db 114674 TCGTATACATCAATATATGATCTATGAACGAAAAACGGGAGGATCTGTTCCTTGGACA 114615
QY 181 CATGGAGAAAAGCAGCACCAGCCAGATGGTCTCCGACGGGGCGACGCTTGCCTTT 240
Db 114615 CATGGAGAAAAGCAGCACCAGCCAGATGGTCTCCGACGGGGCGACGCTTGCCTTT 113475

Db 114614 CATGGAGAAAAGCAGCACCAGCCAGATGGTCTCCGACGGGGCGACGCTTGCCTTT 114555
QY 241 ATTTTCATCGAAGAGCGGATGCGGCACAGCTTTATATCATGAGCACTGAAGCGCGAGAA 300
Db 114554 ATTTTCATCGAAGAGCGGATGCGGCACAGCTTTATATCATGAGCACTGAAGCGCGAGAA 114495
QY 301 GCAAGAAAAGCTGACTGATATCCCATATGGCGTGTCAAAGCGCTATGCTCCCGGACCGT 360
Db 114494 GCAAGAAAAGCTGACTGATATCCCATATGGCGTGTCAAAGCGCTATGCTCCCGGACCGT 114435
QY 361 GAATCGATTCTGCTCACTATCAGTTTGGGAGAGGGGAAAAGCATTTGATGACCGAGAAAA 420
Db 114434 GAATCGATTCTGCTCACTATCAGTTTGGGAGAGGGGAAAAGCATTTGATGACCGAGAAAA 114375
QY 421 ACAGACGAGACAGCTATGAACCTGTTGAAGTGAAGCCCTCTCTTACAAACGGGACCGC 480
Db 114374 ACAGACGAGACAGCTATGAACCTGTTGAAGTGAAGCCCTCTCTTACAAACGGGACCGC 114315
QY 481 AAAGGGCTGACGAGAGGTGCGTATGCCAGCTTGTGCTTGTCAAGCGTAAAGTCGGGTGAG 540
Db 114314 AAAGGGCTGACGAGAGGTGCGTATGCCAGCTTGTGCTTGTCAAGCGTAAAGTCGGGTGAG 114255
QY 541 ATGAAAGAGCTGACAAAGTCAAAAGCTGATCGTCTGCTTTTCTCTGACCGC 600
Db 114254 ATGAAAGAGCTGACAAAGTCAAAAGCTGATCGTCTGCTTTTCTCTGACCGC 114195
QY 601 AAATGGCTGTTTCTCAGCTAATTAACATGAACACATGATGCCAGCAAGCCGATGAT 660
Db 114194 AAATGGCTGTTTCTCAGCTAATTAACATGAACACATGATGCCAGCAAGCCGATGAT 114135
QY 661 GTTTACATAATGTCACCTGGAGTCTGGAGATCTTTAAGCAGGTTACACCTCATCGCGGTCA 720
Db 114134 GTTTACATAATGTCACCTGGAGTCTGGAGATCTTTAAGCAGGTTACACCTCATCGCGGTCA 114075
QY 721 TTGCGATCAAGCTCATTTTCCACGAGGAGGATATCTGCTTGTGTTGGAATGAAAG 780
Db 114074 TTGCGATCAAGCTCATTTTCCACGAGGAGGATATCTGCTTGTGTTGGAATGAAAG 114015
QY 781 GAATATAAGAACTGCTACGCTCTCAAAGCGCTGCTCTATGATATCGAACAGCGCGCTC 840
Db 114014 GAATATAAGAACTGCTACGCTCTCAAAGCGCTGCTCTATGATATCGAACAGCGCGCTC 113955
QY 841 ACATGCTTACTGAGATGCTGGAGTTCATTTAGCGGATGCGCTGATTTGGAGATTCATG 900
Db 113954 ACATGCTTACTGAGATGCTGGAGTTCATTTAGCGGATGCGCTGATTTGGAGATTCATG 113895
QY 901 ATCGGTGGTGTGAAACAGCGCCGATTTGACAAAGACACAGCCAGGGTTTATGTATC 960
Db 113894 ATCGGTGGTGTGAAACAGCGCCGATTTGACAAAGACACAGCCAGGGTTTATGTATC 113835
QY 961 GGCACAGATCAAGCAGTACGGGCACTATTATATTTTCGATTGAAGGCTTGTGTATCCG 1020
Db 113834 GGCACAGATCAAGCAGTACGGGCACTATTATATTTTCGATTGAAGGCTTGTGTATCCG 113775
QY 1021 ATTCGCTCTGAAAAGAGTACATCAATAGCTTTTCTCTTTCACCTGATGAACAGCACTTT 1080
Db 113774 ATTCGCTCTGAAAAGAGTACATCAATAGCTTTTCTCTTTCACCTGATGAACAGCACTTT 113715
QY 1081 ATTGCCAGTGTGACAAAGCGCGACACCGAGTGAAGTTCAGTATCCCGCTTGGACAG 1140
Db 113714 ATTGCCAGTGTGACAAAGCGCGACACCGAGTGAAGTTCAGTATCCCGCTTGGACAG 113655
QY 1141 GAAGAGAAACAGTACTGCGCGAATGACAACTTTGTCTAGGGAGCATACGATATCAATA 1200
Db 113654 GAAGAGAAACAGTACTGCGCGAATGACAACTTTGTCTAGGGAGCATACGATATCAATA 113595
QY 1201 CCGTGAAGAGATTCAATATGCTACAGAAAGCGCGGTGATGTTGAACGCTGGCTGATGAGG 1260
Db 113594 CCGTGAAGAGATTCAATATGCTACAGAAAGCGCGGTGATGTTGAACGCTGGCTGATGAGG 113535
QY 1261 CCGTGAAGAGATTGAGAGTGAACATATCCACTTATTTTAACTATACAGCGGGGTCCG 1320
Db 113534 CCGTGAAGAGATTGAGAGTGAACATATCCACTTATTTTAACTATACAGCGGGGTCCG 113475

QY	1321	CATATGATGTACGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGGGAAGGATAC	1380	Db	1	ATGAAGACGCTGATAACCGGAAAGACCTCATCAAGCTTGTATCCATTACCGATCCCCAG	60	
Db	113474	CATATGATGTACGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGGGAAGGATAC	113415	QY	61	TACGCCCCAGACGGTACCCGCTGCCGATATGTAAAAATCACAAGTAATAACAGAAAGAT	120	
QY	1381	GGGTGCTTTATATCAATCCGAGAGGAGAACCGGCTACGGGCAAGGAATTTGTGAATCGG	1440	Db	61	TATTCGCCGATGTTGCAAAATCGCTACGCTCAAAACAAAGTAACGAAAAACAGGAC	120	
Db	113414	GGGTGCTTTATATCAATCCGAGAGGAGAACCGCTACGGGCAAGGAATTTGTGAATCGG	113355	QY	121	TCGTATACATCAATATATGATCTATGAACGAAACGAGGAGGATCTGTTCTTTGGACA	180	
QY	1441	GTACAGAGGATTTATGGGGAAAGGATTATGACGATGTGATCAGGCTGTGGATGAGGCT	1500	Db	121	TCATATGATTCGCAATATCATGATCTACGCGCGGAAACAAACAGCTTCGGTGCATGGACG	180	
Db	113354	GTACAGAGGATTTATGGGGAAAGGATTATGACGATGTGATCAGGCTGTGGATGAGGCT	113295	QY	181	CATGGAGAAAGCGAAAGCACCGACCCCAAGATGTTCTCCGAGCGGCGCACGCTTGTGCTTT	240	
QY	1501	ATCAACAGAGATCCGCATATTTGATCCTAAGCGGCTCGGTGTACGGCGGGAAGCTACGGA	1560	Db	181	TTTGGAAAGGACGAGAAACAGCATCCCGCTGGTCCGCGCAAAATACCTTGTGATTT	240	
Db	113294	ATCAACAGAGATCCGCATATTTGATCCTAAGCGGCTCGGTGTACGGCGGGAAGCTACGGA	113235	QY	241	ATTTCGTGATCGAAGGCGGATGCGGCACAGCTTTATATCATGAGCACTGAAGCGGAGAA	300	
QY	1561	GGTTTATGACCAACTGGATCTCGGGCAGACGAAACCGCTTTAAAGCTGCCGTTACCCAG	1620	Db	241	ACGTCCCAATCCGCA---GGAAACCGCACAAATTTACGTGATTAAGCGCGCGGGGAGAA	297	
Db	113234	GGTTTATGACCAACTGGATCTCGGGCAGACGAAACCGCTTTAAAGCTGCCGTTACCCAG	113175	QY	301	GCAAGAAAATGACTGATATCCCATATATGGCGTGTCAAAAGCCGCTATGTTCCCGGACGGT	360	
QY	1621	CGCTCGATATCAAAATTTGGATCAGCTTTTCACGGCGTCAAGTATGATCGGCTATTTCTTTACA	1680	Db	298	GCAGAAAAGTACCGGATATTCCTATGATGATACAGCTTGAATGGTCCGCGCAGCGC	357	
Db	113174	CGCTCGATATCAAAATTTGGATCAGCTTTTCACGGCGTCAAGTATGATCGGCTATTTCTTTACA	113115	QY	361	GAATCGATTCTGGTCACTATCAGTTTGGGAGAGGGGAAAGCAATTTGATGACCGGAAAAA	420	
QY	1681	GACTGGCAGCTTGAGCATGACATGTTTGAGGACACAGAAAAGCTCTGGACCGGCTCTCCT	1740	Db	358	AAGTCATTTGCTTTTGTTCAGTGAAGCTGACAAAAGAGGAAGCGTCGACGATGAGAAGAA	417	
Db	113114	GACTGGCAGCTTGAGCATGACATGTTTGAGGACACAGAAAAGCTCTGGACCGGCTCTCCT	113055	QY	421	ACAGACGAGGACAGCTATGAACCTGTTGAAGTGCAGGCGCTCTCTCTACAAACGGGACGGC	480	
QY	1741	TTAAATACGACGAAACGTTGGAGACACCGCTTTTGATGATGCTATGCGAGCGGGATGAC	1800	Db	418	ACCGAGATTTGAAGACCAACCGCGTGAAGTGCATTTCTTTGTCTATATAAAGCAGACGGC	477	
Db	113054	TTAAATACGACGAAACGTTGGAGACACCGCTTTTGATGATGCTATGCGAGCGGGATGAC	112995	QY	481	AAAGGGCTGACGAGAGTGGTATGCCAGCTTGTGCTTGTGCTTGTGTCAGCGTAAAGTCGGGTAG	540	
QY	1801	CGATCCCGATCGACGCGGAGCAGCTGTTTATCGCTCTGAAAAAATGGCGAAGAA	1860	Db	478	CAAGGGTTTAAAGGGGAAAGTATACCGAGCTTGTCTCTGTTTCAAGTGGAAACCGGGGAA	537	
Db	112994	CGATCCCGATCGACGCGGAGCAGCTGTTTATCGCTCTGAAAAAATGGCGAAGAA	112935	QY	541	ATGAAGAGCTGACAAAGTCACAAAAGCTGATGATGTTGATCCTGCTCTTTTCTCTCACGGC	600	
QY	1861	ACCAAGCTTGTCGTTTCCGAAATGCATCGCAAAATTTATCAGCAGCCGAGACCCCAAGA	1920	Db	538	ATGAACAGCTGACCGATCTTGAACGCGACCATTTACGCCATGATTTTACCGTGGGA	597	
Db	112934	ACCAAGCTTGTCGTTTCCGAAATGCATCGCAAAATTTATCAGCAGCCGAGACCCCAAGA	112875	QY	601	AAATGGCTGTTTCTCAGCTAAATTTAACTGAAACAGATGATGCCAGCAGCCGATGAT	660	
QY	1921	CAGCGATCAAGCGCTGAATATATCAGCTCATGTTTGTGATCAACATCTC	1971	Db	598	GATCAGATTCGATTTTTCGCGCAATCAACGGATATGCGGTGT-----GAACGAT	645	
Db	112874	CAGCGATCAAGCGCTGAATATATCAGCTCATGTTTGTGATCAACATCTC	112824	QY	661	GTTCATATAATGCTCACTGGAGTCTGAGATGCTTTAAGCAGGTTTACACCTCATCGCGGTCA	720	
RESULT 3	AX431809	Sequence 224 from Patent WO0229113.		Db	646	GTCTACTTGATGAACCTTTCGACAGTGAATTTGAAGCGCTTCACAGGTTCAAAACGGCAT	705	
LOCUS	AX431809	1958 bp DNA linear PAT 28-JUN-2002		QY	721	TTCCGATCAAGCTCATTTTTCACGACGCGAAGGATATCTTGTCTTGTGGAATGAAAG	780	
DEFINITION	AX431809	Accession		Db	706	TTTTCTTCTACTATCGTTTTTCTCTGACGGCAACACCTTGCAATTTATCGGAACGAGAA	765	
ACCESSION	AX431809	GI:21656613		QY	781	GAATATAAGATGCTACGCTCTCAAGGCGTGGCTCTATGATATCGAACAAAGCGCCCTC	840	
VERSION	AX431809.1			Db	766	GAATATCAAAACGCCACGCTCGATAAAGCATGGCTGATGATATTTGAACAGGAAACATA	825	
KEYWORDS	Bacillus licheniformis.			QY	841	ACATGCTCTTACTGAGATGCTGGACGTTTCATTTAGCGGATGCGCTGATTTGGAGATTCATG	900	
SOURCE	Bacillus licheniformis			Db	826	ACGTGTCTACGGAATGCTTGAAGTCCATCTCAGCGATGCGGTGGCAGGACAGACGCTT	885	
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			QY	901	ATCGGTGGTGTCTCAACAGCGCCGATTTTGGACAAAGGACAGACCAAGGCTTTTATGTCATC	960	
REFERENCE	1	Berka, R. and Clausen, I. G.		Db	886	GTCGGAGGCGTCTTCTCCGAAACCGGCTTGACCAAGACGGAACCGGCTTTTATGTCATC	945	
AUTHORS	Methods for monitoring multiple gene expression			QY	961	GGCAGAGATCAAGGCAAGTACGGGCATCTATTATTTTCGATTTGAAGCGCTTGTGTATCCG	1020	
TITLE	Patent: WO 0229113-A 224 11-APR-2002;			Db	946	GGATCTGACAGGGTCTACAGGCAATTTACTACATATCAATCGAAGGCTTGTGCTTATCCG	1005	
JOURNAL	Novozymes Biotech, Inc. (US); Novozymes A/S (DK)			QY	1021	ATTCGCTCGAAAAAGAGTACATCAATAGCTTTTCTTTTCACTGATGAACAGCACATT	1080	
FEATURES	Location/Qualifiers			Db	1006	GTCCGGTTTGGAAAAAGAGCATGTGAACGGATTCAGCTTGCATCCGAGCAAGAGTGGCTTT	1065	
BASE COUNT	557 a 462 c 518 g 421 t			QY	1081	ATTCGACGTGTGACAAAGCCGACACACCGGAGTGAAGCTTTTACAGTATCC---CGCTTGA	1137	
ORIGIN								
Query Match	41.08;	Score	807.8;	DB	6;	Length	1958;	
Best Local Similarity	64.38;	Pred. No.	1.3e-250;					
Matches 1268;	Conservative	0;	Mismatches	687;	Indels	18;	Gaps	3;
QY	1	ATGAAAAAGCTGATAACCGCAGCAGCATCACAGCGATTGTCTCTGTGACCGATCCTCAA	60					

Db 1066 GCGGCATCCATAGCCCTCGCCGTCGCGGCGAGCGAACTTTACCATCTCCAAATGGGTGAA 1125
QY 1138 CAGGAAGAGAAACAGCTACTGCGCGAATGACAAAGTTTGTTCAGGAGCATACGATATCA 1197
Db 1126 GAAAGAGCGGAGCTACAGATGCCACCAAGCTTTTACGGAAGACATATCATATCG 1185
QY 1198 ATACCTGAGAGATCAATATGCTACGAAGAGCGGCTGATGGTGAAGCGCTGCGTGTATG 1257
Db 1186 GAACCGGAGGAGCTCAATTTTCAGGAGCGGAGCTGACAAATTCACGGATGGCTGATA 1245
QY 1258 AGGCTGACAAATGGAAGGTGAGACACATATCCACATTTCTTAAACATACACGCGGT 1317
Db 1246 AAGCCGCGCAATATGAAAAAGAAACACATCGCTCATCTGGAAGTCACGCGGA 1305
QY 1318 CCGCATATGATGACGACATATATTTTCATGATTTTCAGGTCTGCGCGGCGAAAGGA 1377
Db 1306 CCTCAGCCATGATGCAATGCTTATTTTCATGAATTTCAAGTCTGCGCGCAAGGGA 1365
QY 1378 TACGCGGTGTTTATCAATCCGAGAGAGCGGCTACGCGGCAAGATTTTGTGAAT 1437
Db 1366 AGCGCGCTGATATGTTAAACCGCGGGAAGCGGCTACGCGGCAAGATTTTGTCAAC 1425
QY 1438 CGGTCAGACGAGATTTGGGGAAGAGATTTACGATGTCAGCGGCTGCGGCTGTGATGAG 1497
Db 1426 CGGTGAGGCGGACTACGCGGTGGAGATTTTAAAGATGTAATGGCGCTGTGTGACCA 1485
QY 1498 GCTATCAACAGAGATCCGATATGATCTTAAGCGGCTCGGTGTACGCGGCGGAGCTAC 1557
Db 1486 GTGCTGAACATTTGATTTGCTGATCAAGAAAGGCTCGCATCACCGGAGGAGCTAC 1545
QY 1558 GGAGGTTTATGACCACTGATCGTGGGAGAGCAACCGCTTTAAAGCTGCGGTTACC 1617
Db 1546 GGGGCTTTATGACAACTGGGCTGTGCGGCACACGAAAGCGCTTCAAGCGCGCAGTGACA 1605
QY 1618 CAGCGCTCGATATCAAAATGGATCAGCTTTTACCGGCTGCTGATGATCGCTATTTCTT 1677
Db 1606 CAGAGGTCATTTCCACTGATCAGTTTACGGGTGAGGATGATCGGCTATTTTTC 1665
QY 1678 ACAGCTGGGAGCTGAGCATGATGTTTGGAGACACAGAAAGCTCTGGGACCGGTCT 1737
Db 1666 ACAGCTGGGAGCTGCGGTGAGACCTTTTGAAGATCCAGCAAACTGTGGGACATTCG 1725
QY 1738 CTTTAAATACGCAACGCTGAGACACCGCTTTTGTATGATGATGATGCGGAGCGGAT 1797
Db 1726 CCGCTCAATACGCGGACAGGTGGAGACCCCTTCTCATCTGCTGCGGAGCGGAC 1785
QY 1798 GACGATGCCGATGACGAGCGGAGCAGCTGTTTATCGCTCTGAAAGAAATGGGCAAG 1857
Db 1786 GACAGTGCCGATGAACAGCGGAGACAGCTGTTACGCGGCTGAAAGAAATGGGAAA 1845
QY 1858 GAAACCAAGCTTGTGCTTTTCGATGATGATGATGATGATGATGATGATGATGATGAT 1917
Db 1846 GAATAAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1905
QY 1918 AGACGCGGATCAAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1970
Db 1906 AAGCAGGATGAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1958

RESULT 4

CNSPAX06/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

265118 bp DNA linear BCT 09-SEP-1999
Pyrococcus abyssi complete genome; segment 6/6.
AJ248288 AL096836
AJ248288.1 GI:5458960
complete genome.
Pyrococcus abyssi.
Pyrococcus abyssi.
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
1 (bases 1 to 265118)
Heilig, R.
Pyrococcus abyssi genome sequence: insights into archaeal

chromosome structure and evolution
unpublished
2 (bases 1 to 265118)
Genoscope.
Direct Submission
Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Location/Qualifiers
1..265118
/organism="Pyrococcus abyssi"
/strain="Orsay"
/db_xref="taxon:29292"
164..1615
/gene="bgal"
164..1615
/gene="bgal"
/note="PAB2376"
/codon_start=1
/transl_table=11
/product="BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE)."
/protein_id="CAB50447.1"
/db_xref="GI:5458961"
/translation="MEFYGVVQSAFOEMGDYPRNRDLTRSDMHHWVDRDPVNIKNEL
VSGDLPREGINNYEYEDHRLAKELGINAYOLTIENSRIFPCPIYSEVEVERDGYG
FIKSVKIRKHELEKDELANKREVRHENVLNKLNFTTFTVTLNHTNINLWHDPI
EVRNIEKARAGWDERAIVFESKFAIYAVKEDKYVDFWATFDEPMVTAELGLAP
YVPPGGLNPKAASVIMHIIIAHAYDSIKKPSDKPVGIVILNIIPAYPLNPRDSK
HKAANYDIFHNRLFLVANKGLDIDINGEYKVKHLKNDWIGNNYTYREVVYV
EPKTELPLTFVGVGYSANPNSIPDNNTSDFGWEVYKGLYDSTAEAYEYSE
MVFTENGADSKDILRPYIVDHVKVRLRENGINVLGYFWHALTDNYEWAGFKI
RFLYEVDPITKERIPRRKSVETKYKVKEGIE"
1834..2283
/note="PAB1019"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50448.1"
/db_xref="GI:5458962"
/translation="MPLPDDITFDSVILLKMHSEKRLLEITLAKTIVYLSMTVOR
YLAATLKRNELEFELVDKIYINVPVSEIILIRASKIANLRKGFVLDLEDIITA
VTAETGLLITDEPKYEFPIRRFGLDTPMLDKFLREVELMVEKEIT"
complement(2276..3079)
/note="PAB1341"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50449.1"
/db_xref="GI:5458963"
/translation="MFENLESFEAFKLLLEVLFDLNDPFDKVEKVLCTEPHPDCC
VIGMGITIKLTERGIRIVYACMTDGMGTLDLSLTGHELATIRREERESSKILGVK
KIYLAIVRDTPELVSRVKDLVIRIKKPKDGVFLPDMLPEAHDPHNTGFLALD
AVASPLNPSNVDVEIGLGHQVSYFALYITNKPNTFYDITDMELKRAIKRTHKSO
FPDVMVEWNEPFLRTVALFYKKGKAGVYAEGRIMPGLFYHITPFADLI"
3192..4340
/note="PAB1020"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50450.1"
/db_xref="GI:5458964"
/translation="MKEVSVSSYKEILVKLGLTDETRILEMKGKGIIEFDGIRYVR
FKDSAGLRRGTVIDEYVIGFPHIKRILNRSIGIRIFKRGEFVSEKVGIVYVR
VVMYKGLGITRGGICPFTTPEPQEFKDNPNLILVGMAGPESPVYVEGP
PYVKEDIOFLFDQEIQRSLPVEERLKIAEYGINHVEVFGYTKDDVDLYOLI
ERLSKEGSEIIMKSPDMKIVYTPYANINDIKIGARFYELPGYFTRISRLAF
YIAEKRIKEEFERVAEIGSALLQFVSIFDVEQEDIEHLFKVVRKRIYAKMV
THFEKLGKIEIVDIEIKDWRITFKRLYPDATNEIRELIGGKAFVD"
4351..5658
/note="PAB1021"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

Db	58494	ACTACGGTGAGCCGCAATTATCAAGACCTAATGAGGTAGTTGACGAGGCTTTAAAAACGCT	58435
Qy	1511	ATCCGCATATTGATCTTAAGCGGCTCGGTGTACGGCGGAAGCTACGGAGGTTTATGA	1570
Db	58434	TTGACTTTATAGATGAGGAAGCGTTGGCGTTTACCGGAGGATCATACGGAGGCTTCATGA	58375
Qy	1571	CCAACTGGATCGTCGGCGAGACGACCGGTTTAAAGCTGCCGTTTACCCAGCGCTCGATAT	1630
Db	58374	CAAAATGGATCGTTGGGCACACAAACAGATTCAAAGCTGCCGTAACTCAGCGCTCAATAT	58315
Qy	1631	CAAAATGGATCAGCTTTACGCGCTCAGTGATATCGGCTATTCTCTTACAGACTGGCGAC	1690
Db	58314	CCAACTGGATCAGCTTCTTGGAAACCGGATATCGGCTATTCTCTGCTCCAGATCAGA	58255
Qy	1691	TTGAGCATGACATGTTTGGAGACACAGAAAAGCTCTGGACCGGTCTCTTTAAAAATACG	1750
Db	58254	TGGTGGCATCCCTTTGGATTAACCTAGAAGGCTACTGGGAGAAGGCCATTGAAGTACG	58195
Qy	1751	CAGCAAACTGGAGACACCGCTTTTGTACTGCGATGGCGAGCGGGATGCGCATGCCGA	1810
Db	58194	CTCCAAACGTTGAGACTCGGCTCCGTGATAATTCTACTCCAGCAGACTACCGCTGTGGC	58135
Qy	1811	TGCGAGCGCGGACAGCTGTTTATCGCTCTGAAAAAATGGCGAAGAAACCAAGCTTG	1870
Db	58134	TTCCCGAGGCTTTGCAATTTTACATAGCGTTTGAAGTATTTGGGCAAAACCGTCAGCTGG	58075
Qy	1871	TTCCGTTTTCCGAATGCATCGCACAAATTTATCACCACCGGACACCCAGACAGCGGATCA	1930
Db	58074	CGATATTCACAGTGCAGAAATCATGCTTGAGTAGAGTGGAAGGCCAAAGCATAGGGTTA	58015
Qy	1931	ACGGCGCTGAATTAATATACGCTCATGGTTGA	1961
Db	58014	AGAGGCTCGAGCTCATAGTGGGGGTGGTTGA	57984

RESULT 5	AX041922	265118 bp	DNA	linear	PAT 23-NOV-2000
LOCUS	Sequence	817 from Patent WO0065062.			
DEFINITION	AX041922				
ACCESSION	AX041922.1	GI:11340687			
VERSION					
KEYWORDS					
SOURCE	Pyrococcus abyssi.				
ORGANISM	Pyrococcus abyssi				
	Archaea; Euryarchaeota; Thermococci; Thermococcales;				
	Thermococcaceae; Pyrococcus.				
REFERENCE	1 (bases 1 to 265118)				
AUTHORS	Porter, P., Thierry, J.C., Prieur, D., Dietrich, J., Lecompte, O., Querellou, J., Weissenbach, J., Saurin, W., Heilig, R., Flament, D., Raffin, J.P., Henneke, G., Gueguen, Y., and Rolland, J.L.				
TITLE	Genome sequence and polypeptides of pyrococcus _1 (abyssi), fragment and uses thereof				
JOURNAL	Patent: WO 0065062-A 817 02-NOV-2000;				
	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR); IFREMER				
	INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)				
FEATURES	Location/Qualifiers				
source	1..265118				

Query Match 12.1%; Score 238.2; DB 6; Length 265118;
Best Local Similarity 55.3%; Pred. No. 1.4e-64;
Matches 482; Conservative 0; Mismatches 386; Indels 3;

RESULT 6	
AR116209	
LOCUS	1896 bp DNA
DEFINITION	Sequence 2 from patent US 6133012.
ACCESSION	AR116209
VERSION	AR116209.1 GI:14096531
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown..
REFERENCE	Unclassified. 1 (bases 1 to 1896)

The other authors are at the National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan. All the sequence with length 100 codons or more between ATG or GTG and stop codon are defined as CDS. Homology analysis is performed by Smith-Waterman algorithm against GenBank and GenPept release 103; EMBL release 52.0; SwissProt release 34.0; PIR-protein release 54.0; and OWL release 29.5. E-mail address for comments and questions: genome@nitech.go.jp. Restriction map, ORF organization, sequence alignment and more information are available at w.w.w. site of Biotechnology Center, URL: <http://www.bio.nite.go.jp/>.

Location/Qualifiers
1. .233000

source

/organism="Pyrococcus horikoshii"
/strain="OT3"
/db_xref="taxon:53953"
/map="544001-777000 nt. position"
complement(645..1067)
/gene="PH0605"
complement(645..1067)
/gene="PH0605"
/note="similar to PIR:A64312 percent identity: 51.111 in 136aa; PIR:A64312 percent identity: 51.111 in 136aa; Swiss_Prot:P41375 percent identity: 44.762 in 106aa"
/codon_start=1
/transl_table=11
/product="140aa long hypothetical translation initiation factor eIF-2 beta"
/protein_id="BAA29694.1"
/db_xref="GI:3257011"
/translation="MEIDYDYKLEKAYEELPENVKHKSFRFPGALVTIEGNTITENKDIKIAALNRPDHLKLEIATAGTLEGRVVGLOGRFTPLYIANKLKYYIK
EVICPVGSPDTKIIKRDREYFLKCEACGAETPIQHL"
complement(1110..4436)
/gene="PH0606"
complement(1110..4436)
/gene="PH0606"
/note="similar to Swiss_Prot:P30665 percent identity: 41.538 in 331aa. motif-ATP/GTP-binding site motif A (P-loop); protein splicing signature"
/codon_start=1
/transl_table=11
/product="1108aa long hypothetical cell division control protein"
/protein_id="BAA29695.1"
/db_xref="GI:3257012"
/translation="MERELIERFLKFFREYAEAGEEPLYMGKIKOLLITPKRSIVI
NMHLNSYDPELAEVIEPEECILAEADAIQILKEDFLREDVPIKHAFHNLPTKL
MYKEVGAEHINKLIQVEYVTRVTEIKPFVAVFVCKDCGHEMVPQYEGFVAVK
KCEQSGKNVQDVEKSFVNFQMFRIQDRPSLGGQMPFIDGILLDDIVDTAMPK
DRVIVGLRVIOEKREKVPFVKVIEVNHIEPISKELELEISPEDEQKIRELAKRK
DIVSAIVDSIAPALYGYKEVKKGIALALFGYSRTLPDCTLRGLDHLVLLGDPGVAK
CVDYTELVDGGRKKGIEVEEAKKAEKGLGRVDDGGFYAPINLELYALDVRTL
KVRKQADIAKMTTPKRLIRKRGREIRVTPHFFTEEGRIKTKAYELKVE
KIATPREAPEAEIWEVEYEEKPNNSWYDQVPEHNFANGIFVHNSQLRY
VANAPRAIYTSKSSAAGLCVAPDITLNTONGRVEIGKFEVEMKVEGEISEEGIS
YAPCRFTKTFKDGIVESPVRVVKLRAPKLVIRKTKENGSIATLTRETLLITINDG
ELSWVEAGEVKGTVGVSEKDVIPGAGKTIRDSKLYNMEMEKVLYLTREVRKA
IEKLEIMNPNKIPGVQSEYELLRLKLTNDRVRNETLILLSDVSDAHELAKRK
IEKIEINSEVHVEKVEVGEVGYEYDITVEGSHNFVANGFIHNTAAVVRDEF
TGGWYELAGLVADGVALIDELKMSDRDSVTHALEQQOTISLKGAGITATLNAR
TTVAANAPKGRFNRMRVSEQIDLPPTLLSRFDLIFIMDEPKDVESEIARHLIK
VRRGEVVTPIKIPHELLRKIYAYARKNIHPVISEAMEEIEKYYVKMRSAKSSGD
ETKPIPTAROLEALIRLSEAHMRSLSPITREDAREAKLMEYTLKQIATVDEAGOL
DYTILEVQSQARKLSKVERILIVKELRTSEKGAIDIDLEAAKFGTEKSEAREIL
EKLLKGGQIYMPENGYKTV"
1630..1992
/gene="PH0607"
1630..1992
/gene="PH0607"
/note="motif-glutamine amidotransferases class-II active site"
/codon_start=1

gene

CDS

FEATURES

1619 AGCGCTCGATATCAAAATGGATCAGCTTTACGCGCGTCAGTGATATCGGCTATTCTTTA 1678
1619 AGCGCTCGATATCAAAATGGATCAGCTTTACGCGCGTCAGTGATATCGGCTATTCTTTA 1678
1538 AGAGATCAAAATTTCAAAATGGATAGCTTTCGGGACACGATATAGGTATTACTTTG 1597
1538 AGAGATCAAAATTTCAAAATGGATAGCTTTCGGGACACGATATAGGTATTACTTTG 1597
1679 CAGACTGGCAGCTGGACATGACATGTTTGGAGACACAGAAAGCTCTGGACGGGTCTC 1738
1679 CAGACTGGCAGCTGGACATGACATGTTTGGAGACACAGAAAGCTCTGGACGGGTCTC 1738
1598 CTCAGATCAAAATAGGAAAAGATCCCTGGAGCAACTTTGGAGGTTATTGGGAAAAGAGCC 1657
1598 CTCAGATCAAAATAGGAAAAGATCCCTGGAGCAACTTTGGAGGTTATTGGGAAAAGAGCC 1657
1739 CTTTAAATACGACGACAAAGCTGGAGACACCGCTTTTGATACCTCATCGGACGGGATG 1798
1739 CTTTAAATACGACGACAAAGCTGGAGACACCGCTTTTGATACCTCATCGGACGGGATG 1798
1658 CATTAAAGTACGCTCCCAAGCTTGAACCTCCCTGCTTATAATCACTCTACCGAAGACT 1717
1658 CATTAAAGTACGCTCCCAAGCTTGAACCTCCCTGCTTATAATCACTCTACCGAAGACT 1717
1799 ACCGATCCCGATCGAGCAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATGGGCAAGG 1858
1799 ACCGATCCCGATCGAGCAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATGGGCAAGG 1858
1718 ACAGGTGTTGCTTCCCGAGGCATTCGAACCTTCATATCCCTAAATACCTGGGGAAGA 1777
1718 ACAGGTGTTGCTTCCCGAGGCATTCGAACCTTCATATCCCTAAATACCTGGGGAAGA 1777
1859 AAACCAAGCTGTCGGTTTCCGAATGATCGACAAATATATCATCGACACCGACACCCAA 1918
1859 AAACCAAGCTGTCGGTTTCCGAATGATCGACAAATATATCATCGACACCGACACCCAA 1918
1778 GAGTTGAATGGCAATATTCACAGGAGAAAATCATGACCTAAGTAGATCTGGGAAGCAA 1837
1778 GAGTTGAATGGCAATATTCACAGGAGAAAATCATGACCTAAGTAGATCTGGGAAGCAA 1837
1919 GACAGCGGATCAAGCGCTCAATATATCATGCTCATGTTTGATCAACATCT 1970
1919 GACAGCGGATCAAGCGCTCAATATATCATGCTCATGTTTGATCAACATCT 1970
1838 AGCACAGGTTAAAGACTTGAACCTAATAGCAGGATGGTGAAGAAATGGCT 1889
1838 AGCACAGGTTAAAGACTTGAACCTAATAGCAGGATGGTGAAGAAATGGCT 1889

RESULT 8

AP000003

LOCUS

DEFINITION

Pyrococcus horikoshii OT3 genomic DNA, 544001-777000 nt. position

(3/70).

ACCESSION

AB009484 AB009485 AB009486 AB009487 AB009488 AB009489

AB009490 AB009491 AB009492 AB009493 AB009494 BA000001

AP000003.1 GI:3236130

VERSION

KEYWORDS

SOURCE

ORGANISM

Pyrococcus horikoshii (strain:OT3) DNA.

Pyrococcus horikoshii

Archaea; Euryarchaeota; Thermococci; Thermococcales;

Thermococcaceae; Pyrococcus.

1 (sites)

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Okura, R., Nakazawa, H., Takamiya, M., Ohnaka, Y., Funahashi, T., Tanaka, T., Kudo, Y., Yamazaki, J., Kishida, N., Oguchi, A., Aoki, K., Nakamura, Y., Robb, T. F., Horikoshi, K., Masuchi, Y., Shizuya, H., and Kikuchi, H.

Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3

DNA Res. 5 (2), 55-76 (1998)

98344137

2 (bases 1 to 233000)

Tanaka, T., Kawarabayasi, Y. and Kikuchi, H.

Direct Submission

Submitted (11-JUN-1998) Yutaka Kawarabayasi, National Institute of Technology and Evaluation, Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan

(E-mail: genome@nitech.go.jp, Tel.: +81-3-3481-8951, Fax: +81-3-3481-8424)

On or before Jul 28, 1998 this sequence version replaced

gi:3130861, gi:3130869, gi:3130912, gi:3130938, gi:3130941,

gi:3130952, gi:3130999, gi:3131044, gi:3131081, gi:3131085.

AB009484-AB009494: submitted (10-DEC-1997)

Kawarabayasi, Y. is officially affiliated with the National Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki

305-0046, Japan.

Robb, T. F. is at the Center of Marine Biotechnology, University of Maryland, Baltimore, MD, USA.

Horikoshi, K. is at the Japan Marine Science and Technology Center, Yokosuka, Kanagawa 237-0061, Japan.

Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041, Japan.

Shizuya, H. is at the California Institute of Technology, Pasadena, CA, USA.

COMMENT

gene

CDS

LOCUS AE013176 12882 bp DNA linear BCT 09-MAY-2002
DEFINITION Thermoanaerobacter tengcongensis strain MB4T, section 203 of 244 of the complete genome.
ACCESSION AE013176 AE008691
VERSION AE013176.1 GI:20517339
KEYWORDS
SOURCE Thermoanaerobacter tengcongensis.
ORGANISM Thermoanaerobacter tengcongensis
Bacteria: Firmicutes; Clostridia; Thermoanaerobacteriales;
Thermoanaerobacteraceae; Thermoanaerobacter.
REFERENCE 1 (bases 1 to 12882)
AUTHORS Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J. and Yang, H.
TITLE A Complete Sequence of the T. tengcongensis Genome
JOURNAL Genome Res. 12 (5), 689-700 (2002)
21992816
PUBMED 11997336
REFERENCE 2 (bases 1 to 12882)
AUTHORS Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Genetics and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China
3 (bases 1 to 12882)
AUTHORS Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
4 (bases 1 to 12882)
AUTHORS Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China
FEATURES
source
1. 12882
/organism="Thermoanaerobacter tengcongensis"
/strain="WB4T"
/db_xref="taxon:119072"
terminator complement(70..90)
gene complement(93..1765)
/gene="TTE2335"
CDS complement(93..1754)
/gene="TTE2335"
/note="Best Blastp hit = gi|10175658|dbj|BAB06755.1| (AF001517) two-component sensor histidine kinase [Bacillus halodurans], score 177, E-value 4.00E-43"
/codon_start=1
/transl_table=1
/product="Diguanylate cyclase/phosphodiesterase domain 1 (GGDEF)"
/protein_id="AAM25476.1"
/db_xref="GI:20517340"
/translation="MGKRYELFNTALATGGLVAFALXHYGLGIDFRFLILLFLA IILDNGIMTYDKLSLSPGSGVAAFLIFGTGGAATLMTVSMFDTLVMRRKIKNGFL NGAMFALYLPACGLVPEFMGKIGIEISIAOVKYLTVVIVISFLNNELTYALAKOGL ILPEKWNESVLLELTYVLMVPAAMLLAVYFKHSLMFEVLSLTPLVFAYVFMVR VLVKANKRLNATYEMKMNKSLDELQILDTIVEVISQVAVSAAALYLDTSNGIATL KVAGKGRGFRSEYFKGEGGLIGKVISSNKVIAKLNKEDRSIDREKIFANYGLSID APLRSSGVPICGLLIHLVETNAFDNDDSVRIEIIIVDQASVAITNAKKIYEVTRKSID PLTKTNNRYFNDALENIMRADENEPVSLIMFDLNFQINDTYGHLIGDEVLKEV ARRIKNNVRSNDIVARFGGEFAVILPKLTAEOYMTAERIRTVSSKPIKTEKGDIY VTITGGVADYPSKADSAEKLVSHADRALYAGGSKGRNKAITYEV"
misc_feature complement(114..611)
/gene="TTE2335"
/note="Pfam match to entry GGDEF, GGDEF domain, score 212, E-value 9.30E-60"
misc_feature complement(612..1031)
/gene="TTE2335"
/note="Pfam match to entry GAF, GAF domain, score 59.7, E-value 6.40E-14"

RBS complement(1761..1765)
/gene="TTE2335"
terminator complement(1912..1934)
/note="putative rho-independent transcription terminator"
gene complement(1941..4016)
/gene="NrdA"
/note="TTE2336"
CDS complement(1941..4016)
/gene="NrdA"
/note="Best Blastp hit = gi|11499254|ref|NP_070492.1| ribonucleotide reductase (nrd) [Archaeoglobus fulgidus] gi|17483938|pir|G69457 ribonucleoside-diphosphate reductase (EC 1.17.4.1) large chain nrd - Archaeoglobus fulgidus gi|2648891|gb|AA89584.1| (AE000988)
ribonucleotide reductase (nrd) [Archaeoglobus fulgidus], score 440, E-value 1.00E-122"
/codon_start=1
/transl_table=1
/product="Ribonucleotide reductase alpha subunit"
/protein_id="AAM25477.1"
/db_xref="GI:20517341"
/translation="MFKRVAKATATKEEMFYKVMSEWKFIPGGRILAGAGTGRGV TYNCFVIVPEANDPKGNDRAAIMDTIAKMLEINSRGGGVGINWTLRPRGAYVKV VGGTSSGAYWMLAANEVATQVGGSRRAALMFLMDWHPDIEEFIKVKDLTKMQQ ANLSVADAFMEAVQDKWKWLEFPDTSHPNDVDAEWKGLNEMKAGYPTVYKEIP ARELNKIVTAAWESAEPGVFLERYNKLSTVYITKLIITNPGELGPEYGVNCLG AINLVAFVNGKINFEELETYKIVRFLDNLGAYVLQNKEMAQKLRVGLGLM GLADAILMNRYSSEESUKATEEIVRRIRDAAYRASVLAKEKGPPEFVADKYLKG QTIQPKDIDRDIKAYIRNATILTQPTGTTSILAGSSGIEPNFAKEYVRKDRGTG THTVHKMLADYPAFVAHEVTPPEHVQMVAQYKIDSSISKTINLPRATVDDIDNI YLTFDNGKPFVFNIGRAGSDVAAFEARLISLALRGVAVEDIIDQLVGIGGA TTIGGGRNVLSPDDAIGKALNRNGYEQGNSVKVSATNVDICPCGMATFYTGGCKT CLNCGYSQC"
misc_feature complement(2442..3872)
/gene="NrdA"
/note="Pfam match to entry ribonuc_red_lgC, Ribonucleotide reductase, barrel domain, score 126.6, E-value 4.60E-34"
gene complement(4033..4224)
/gene="TTE2337"
CDS complement(4033..4209)
/gene="TTE2337"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM25478.1"
/db_xref="GI:20517342"
/translation="MHLKLEKNTKCIIEDVDIONIVFYVTKGKINASSVERTEGEGF SGQVCLKRGQGR"
RBS complement(4220..4224)
gene complement(4297..4808)
CDS complement(4297..4794)
/gene="TTE2338"
/note="Best Blastp hit = gi|11349142|pir||A83493 hypothetical protein PA1234 [imported] - Pseudomonas aeruginosa (strain PA01)
gi|1947163|gb|AA04623.1|AE004552.13 (AE004552)
Hypothetical protein [Pseudomonas aeruginosa], score 55.8, E-value 3.00E-07"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAM25479.1"
/db_xref="GI:20517343"
/translation="MSLKDLFEKGLTFDEFIETSEDITRQIRERYDKTVLEKDVQK KIVSLPKVNIILAFATETPCDSQVNPVIAKIASYNKTFELKIVKREGNEEYMKPYVE GKPRIPTFVMDENEFKFWHVERPOIVRELEKKEVADWKRDLKGYDSEIIRIIVY IFLRP"
RBS complement(4804..4808)
gene complement(4809..6244)

	CDS	/gene="ArgD" /note="TTE2339" complement(4809..6230) /gene="ArgD" /note="Best Blastp hit = g1 12517634 gb AAG58206.1 AE005536_7 (AE005536) probable ornithine aminotransferase [Escherichia coli O157:H7 EDL933] g1 13363428 dbj BAB37378.1 (AP002564) probable ornithine aminotransferase [Escherichia coli O157:H7], score 302, E-value 5.00E-81" /codon_start=1 /transl_table=11 /product="PLP-dependent aminotransferases" /protein_id="AAAM25480.1" /db_xref="GI:2051734"
	misc_feature	/translation="MDEKVKIKPEIPEYITLEBALTLNKNVKNKYINSVFVSM SMLQDFKLFPRAKGSVDSENEYDFLGYGALNGHNPDEVIEAVPKVDMPNML QASIGNLPGLVAHLNLRVTGNLKRSFFCNSGAEVEGALKLAKIASGKKKIYVCENS FHGSMGALSYPGRKRYOKYPELVIPETVAVPFGDEKALEELKEKDVAAFIVEPIQG EGGVIVPEGYRRARELCTKYEAYLVIDEIQTGFRTKMFACEHEEVDPDIATLAK SLGGVMPICAGYITTDWOKAYGTMEKALLHTSTFGNTYACAAAIASTQAIIEKL SEAAKEGEVFLGRILKELKHHPKLIKDVRKGILLIGEPNOPEGGLDKLSGAISK LSSEYIGSLAAELQNKRHIITAYTLNNPNVIRLEPLIVTKEQIDKVVDALDEILTR NSSFLGITVSQAIVGLSGLFKQ" complement(4875..6071) /gene="ArgD"
	RBS	/note="Pfam match to entry aminotran_3, Aminotransferase class-III, score 536.6, E-value 8.80E-161" complement(6240..6244)
	gene	/gene="ArgD" complement(6248..6466) /gene="TTE2340"
	CDS	complement(6248..6454) /gene="TTE2340" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAAM25481.1" /db_xref="GI:20517345" /translation="MITKEMIDRLNELYHKSKBSGLTEERKEQRRLREAYVKEIKER VKRELNLFDASHHHHCJHHHEH" complement(6462..6466) /gene="TTE2340" 6648..7300 /gene="TTE2341" 6648..6652 /gene="TTE2341" 6662..7300 /gene="TTE2341" /note="Best Blastp hit = g1 11497972 ref NP_069196.1 conserved hypothetical protein [Archaeoglobus fulgidus] g1 7483000 pir H69294 conserved hypothetical AF0360 - Archaeoglobus fulgidus g1 2650269 gb AAB50871.1 (AE001079) conserved hypothetical protein [Archaeoglobus
	Query Match	11.1%; Score 218.2; DB 1; Length 12882; Best Local Similarity 54.8%; Pred. No. 1.8e-58;
	Matches	455; Conservative 0; Mismatches 373; Indels 3; Gaps 1;
QY	1140	GGAAGAAACAGTGACTGGCGGAATGACAAGTTTGTCTCAGGAGCATGCATATCAAT 1199
Db	8176	GCAGGAGATACAGATTACTGACTTTTAATGAGTGGTGCAGAGAAGAAAGAACTTTCAA 8117
QY	1200	ACCTGAAGATTCAATATGCTACAGAGACGGCGTGTGTCAGCGCTGGCTGTAGTAG 1259
Db	8116	GCCTGAAAGAGTAGAGGTAGACAGCGCCAGGGCATTTTATAGATGGTTGGGTATAAA 8057
QY	1260	GCCTGCACAAATGGAAGGTGAGACAACATATCCCACTATTCTTAACATACACGCCGTCC 1319
Db	8056	GCCTGTAGACTATGAAGAGGGAAGCAAGTATCCCTCCAATATATAGATATACATGTGTGTC 7997
QY	1320	GCATATGATGTACGGACATACATATTTTCATGAGTTTCATGCTGCGCGCGAAGGATA 1379
Db	7996	TAAAAACAGATATATGTTGAAGTATACTTCCATGAATGCAGTACTGGCTTCAAGAGGTA 7939
QY	1380	COCGGTCGTTTTATATCAATCCGAGAGAAACACCGCTACGGGCAGAAATTTGTGATGC 1439
Db	7936	TTTTGTGTTTTTCTGCATCCGTTAGGAAGCATGGCAGAGAACCAATTTG---CTGA 7880
QY	1440	GGTACAGAGAGATATGGGGAAAGGATATGACAGTGTGATGTCAGGCTGTGGATGAGGC 1499
Db	7879	TATAAGAGGAAATATGGACAATTGACTATCAAGATATAATGAAATTTACAGATTATGT 7820
QY	1500	TATCAACAGAGATCCGCATATTGATCCCTAAGCGCTCGTGTACGGCGGAAGCTACGG 1559
Db	7819	GCTTGAATAATTACAAGACATTTGACCCGAAAGAGTTGGGGTTTACAGTGGATCTTACGG 7760
QY	1560	AGGTTTTATGACCAACTGGATCGTGGGCAGACGAACCGCTTTTAAAGCTGCGCTTACC 1619
Db	7759	TGGATTTATGACAAACTGATAATAGGCATACAGATAGGTTTAAAGCAGCAGTATCTCA 7700
QY	1620	GGCTCGATATCAAAATTTGGATCAGCTTTTACGGCGTCTAGTGATATCGCTATTTCTTAC 1679
Db	7699	AAGGAGCATATCAAACTCGTTTACAGAGTTTGGAACTACAGATATAGTATTACTTTGT 7640
QY	1680	AGACTGGCAGCTTGAGCATGACATGTTTTGAGGACACAGAAAGCTCTGGGACCGTCTCC 1739
Db	7639	GCAGATCAGGTGGCGGTACTCCCTGGGACAACTTTTGAGAAATACTGGGCAATTTCTCC 7580
QY	1740	TTTAAATACGAGCAAAACGTTGGAGCACACCGCTTTTGATCTGCATGGCAGCGGATGA 1799
Db	7579	GCTTAAATATGCTGATAAAGTGAAGACTCCGACATGTTCTCGCATCTGATGAGGATTT 7520
QY	1800	CGATGCCCGATCGAGAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATTTGGCAGGA 1859
Db	7519	TAGGTGCTGGCTTGTGAGGCTCTTCAAATGTTTCAAGCTTTCACGCCCTTAACTTGGTGTA 7460
QY	1860	ACCAGCTTGTCGGTTTTCCGAATGCATCCACAATTTATACGCACCGGACACCAAG 1919
Db	7459	GTCAGGTTGGTGATCTGTGATGGTGAACACCGACCTTCAAGGAGCGGAAACCAA 7400
QY	1920	ACAGCGGATCAAGCGCCCTCAATATATCATGCTCATGCTGTTTGTCAACATCT 1970
Db	7399	GCACAGGATAAGAGGCTTAAAGAGATAACTGATTTGGTTTAAAGTACCT 7349
	RESULT 10	
	LOCUS	AX436451 477 bp DNA linear PAT 28-JUN-2002
	DEFINITION	Sequence 4866 from Patent WO0229113.
	ACCESSION	AX436451
	VERSION	AX436451.1 GI:21661259
	KEYWORDS	.
	SOURCE	Bacillus clausii.
	ORGANISM	Bacillus clausii
	REFERENCE	1 Berka,R. and Clausen,I.G. Methods for monitoring multiple gene expression Patent: WO 0229113-A 4866 11-APR-2002; Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
	FEATURES	Location/Qualifiers source 1..477 /organism="Bacillus clausii" /db_xref="taxon:79880" BASE COUNT 138 a 89 c 118 g 132 t ORIGIN
	Query Match	9.8%; Score 192.2; DB 6; Length 477; Best Local Similar

SOURCE
Aeropyrum pernix
ORGANISM
Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococcaeae; Aeropyrum.
REFERENCE
AUTHORS
1 (sites)
Kawarabayashi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y.,
Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H.,
Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H.,
Takamiya, M., Masuda, S., Funahashi, T., Tanaka, F., Kudoh, Y.,
Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K.,
Nakamura, Y., Nomura, N., Sako, Y. and Kikuchi, H.
TITLE
Complete genome sequence of an aerobic hyper-thermophilic
crenarchaeon, Aeropyrum pernix K1
JOURNAL
DNA Res. 6 (2), 83-101 (1999)
MEDLINE
99310339
REFERENCE
2 (bases 1 to 251700)
Tanaka, T., Hino, Y., Kawarabayashi, Y. and Kikuchi, H.
AUTHORS
Direct Submission
TITLE
Submitted (14-DEC-1998) Yutaka Kawarabayashi, National Institute of
JOURNAL
Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10,
Shibuya-ku, Tokyo 151-0066, Japan [E-mail: kyutaka@kazusa.or.jp,
Tel.: +81-3-3481-8951, Fax: +81-3-3481-8424]
COMMENT
Kawarabayashi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Nomura, N. and Sako, Y. are at Kyoto University, Sakyo-ku, Kyoto
606-8502, Japan
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 109; EMBL release 56.0; Swissprot
release 36.0; PIR-protein release 57.0; and OWL release 31.0.
E-mail address for comments and questions: genome@nite.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: <http://www.miln.nite.go.jp/>.
FEATURES
source
1..251700
/organism="Aeropyrum pernix"
/strain="K1"
/db_xref="taxon:56636"
complement(948..1026)
/gene="APETRNA28"
complement(948..1026)
/gene="APETRNA28"
/product="tRNA-Gly"
/notes="codon recognized: GGG; anticodon: CCC"
/anticodon="(pos:988..990,aa:Gly)
1157..2149
/gene="APE1559"
1157..2149
/gene="APE1559"
/notes="similar to PIR:A69437 percent identity:31.228 in
285aa."
/codon_start=1
/transl_table=11
/product="330aa long hypothetical protein"
/protein_id="BAA80558.1"
/db_xref="GI:5105245"
/translation="MSGPGTYSPPSRLLQYLCCASVAGARLVPGLGGRFPFNCVL
VEDAGFVLVDTCFPERGLVDVVFVTFPHDHIRGFRHLRAGRVAAGPEEPYA
GLRLALRFAGERWEISMARLVMDEGVPEGPYPRPDLCLFRGVCIKTYIPARG
HLQHTLLPGGLVHLVDIDLTGFGPWYGNPEASPELFLDIEAAEAEAGCYLSAH
KDRVYSRGEGVLAARFAGRLVEOAGVYKALEAAGRLPWPWELTGRGIIYRRYLEGM
EALMSYFPGVMIEKLLGLLGLWGCTRKTPGGIVARDCHLIGVTKDKTRNIUSLA"
complement(1196..1684)
/gene="APE1560"
complement(1196..1684)
/gene="APE1560"
/codon_start=1
/transl_table=11
/product="162aa long hypothetical protein"
/protein_id="BAA80559.1"
/db_xref="GI:5105246"
/translation="MDKPPGQEQGVLEVPSPGVLNADPPEAEVLPRLPRLRH
AFQIHNPASHADPLPLPLAGKPEKPLKPPSIGLLTLRGSHSPASQPMPDPDWVVEV
SEHHYINHQPTLPGLTPRIHENLEPRILNQHTVAKGKPPPKRPNKPSHPSTAPQ
IL"
2258..2437
/gene="APES056"
2258..2437
/gene="APES056"
/notes="similar to OWL:SOU61050 percent identity:41.667 in
24aa."
/codon_start=1
/transl_table=11
/product="59aa long hypothetical protein"
/protein_id="BAA80560.1"
/db_xref="GI:5105247"
/translation="MARVRIVAKNGPILLEVKGVTALCRGHSNNKPKYCDGTHAR
VGRQADEKVAFAEF"
2536..2952
/gene="APE1561"
2536..2952
/gene="APE1561"
/notes="motif=neutral zinc metalloproteinases, zinc-binding
region signature"
/codon_start=1
/transl_table=11
/product="138aa long hypothetical protein"
/protein_id="BAA80561.1"
/db_xref="GI:5105248"
/translation="MRVERAPDYCRAAEVLYDALGDYIDVSRVYCVSRVSGSRGRAYA
RYGTGCAAAALGPNPVIIEVSEKFDPLPEEKVKVVIHELLHPIRTFTSGGLRAH
GPLVNGRRVAALYRRVGRDARVMGILREALDGGPQE"
complement(2599..3063)
/gene="APE1562"
complement(2599..3063)
/gene="APE1562"
/codon_start=1
/transl_table=11
/product="154aa long hypothetical protein"
/protein_id="BAA80562.1"
/db_xref="GI:5105249"
/translation="MGSRVSWPGRARPCYAGGLGSPFSPWPPPCWELFLRASI
OGFAEDAHNPGITPDPPVORCNPPVYQCGAMGPEAAEGAGVVEQLMYDNLNLLQWQ
RVELLDDLNYNAGVIAESRSPSPRPVDPRIPTTPPPAPHAVHFGDVIDV"
complement(2906..3376)
/gene="APE1563"
complement(2906..3376)
/gene="APE1563"
/codon_start=1
/transl_table=11
/product="156aa long hypothetical protein"
/protein_id="BAA80563.1"
/db_xref="GI:5105250"
/translation="MEKEDGVLRLRTTPPKGCDLTPDVHAYLGSLGVREIALGYTRV
YCRDGAEEVLLDLEGMLTIAOPPEPPIAAVTLSPGNAILLITLIERODGATAASTA
KGDAGEPGVVARREGAGLRWGLGLKAILAMAGAAALLGAILLEGLHPRLRGGCP"
complement(3615..4268)
/gene="APE1564"
complement(3615..4268)
/gene="APE1564"
/notes="similar to PIR:H70375 percent identity:45.455 in
110aa, motif=C-5 cytosine-specific DNA methylases
signatures"
/codon_start=1
/transl_table=11
/product="217aa long hypothetical protein"
/protein_id="BAA80564.1"
/db_xref="GI:5105251"
/translation="MAGVGRPPPPSVGSVVVADTMTDAKLREVLGARSLEAARIM
ARLARSRMGLTVITTSVRAEIERFMIGNGVHPTTIARLMAWVRVKPTTHLRSLP
AAMFRYVETVARLDKGLRVAEDHVRRLALRSRDEGDIVRSRLREKYRTRKGMDS


```
VEDVDTLLALELGAIAVTSDEGLRRAEDLGLTLVLPVELLEY ILALEEEVKVEE"
3888. .4304
/gene="APE1565"
3888. .4304
/gene="APE1565"
/codon_start=1
/transl_table=11
/product="138aa long hypothetical protein"
/db_xref="GI:5105252"
/translation="MILSHPKLVEPGLNGLHIPSHGWEAQLMRGRLNPHPRHKP
GNGCRHVAHYPEHYLYSDARSYVHRPHPGSPGOYPSHYPSGLLQAPGPEDLAQLG
VQKGRVCHHNAHARRRPPAHASHPLQPPCSLLV"
3955. .4299
/gene="APE1566"
3955. .4299
/gene="APE1566"
/note="motif=prokaryotic membrane lipoprotein lipid
attachment site"
/codon_start=1
/transl_table=11
/product="114aa long hypothetical protein"
/protein_id="BAA80566.1"
/db_xref="GI:5105253"
/translation="NAAGRRSCVVGGLTLTHAISLAWVVGCTPPSPMNLISALTLG
VMYVSPILLEARASILAIILAASSRLAPRTSRSLASVMNVVSATITLTLGGGLPPT
PATPSNPHVPHP"
complement(4296..5405)
/gene="APE1567"
complement(4296..5405)
/gene="APE1567"
/note="similar to Swiss_Prot:Q57857 percent
identity:33.430 in 344aa. motif=aminoacyl-transfer RNA
synthetases class-II signatures"
/codon_start=1
/transl_table=11
/product="369aa long hypothetical protein"
/protein_id="BAA80567.1"
/db_xref="GI:5105254"
/translation="MDPGSVDEALSTRSVRSVRVEDVYVYFRQRPQVPEGLTIVA
FRGQIVVMGYPSTKRLMLPRVPPFPPTVLVEERKMNGYNRVFTLGMVYAAAT
RGGLTCTYRRLRLYGGALQKLELDGAEGSFAGEVVGLENPYTRYIYBEAPGFG
YFIFDIKGGROLPRVKFSLAPEKLVNLLAEIPATASGVERLYTIVDELEKRGK
EGVIVDPGRVPELKYTSRTNIGDIRLGMVPPEGRSFLFPRLIRIFREWEYGR
RYGELGEAILAPATEAVEAVSRGRLVEFEELVANEVEAEVYAFSLGVHLEIA
GVARGVDGVRAFRPKSEGIARILEITGISPLD"
complement(5525..5854)
/gene="APE1568"
complement(5525..5854)

Query Match
Best Local Similarity 8.2%; Score 161.8; DB 1; Length 251700;
Matches 369; Conservative 0; Mismatches 302; Indels 6; Gaps 2;

QY 1297 ATTCTTAACATACACGGCGTCCGCATATGATGATGACGACATACATATTTTCATGAGTTT 1356
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172547 ATTCTTATATACATGGGGCGCTAAGACGAGCTACGCGTACGCTTCATACACAGTTTC 172606

QY 1357 CAGGTGCTGGCGGGAAGATACGCGGTCTGTTTATATCAATCCGAGAGAACCCACGGC 1416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172607 CAAATGCTACGCTCTCAGGGCTTTGCGGTGATTTACTCGAACCCCGGGGAGTACGGT 172666

QY 1417 TACGGCAGCAATTTGTGATGGCTCAGAGGAGATTATGGGGGAAGGATTATGACGAT 1476
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172667 TACAGCGAGGTTGCCCAT --- CTAAGGGGTAGATACGGGTTCACGATTACTCTGAG 172723

QY 1477 GTGATCAGGCTGTGATGAGGCTATCAAAACGAGATCCGCATATTGATCCTTAAGCGGCTC 1536
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172724 CTTATGAAGGTTGTTACACAGGCCCTCTCGGACTTCCACACAGCTTGACCCGACAGGGGT 172783

QY 1537 GGTGTACCGGGCGGAAGCTACGAGAGTGTATGACCAACTGGATCGTCGGGACGAGCAAC 1596
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172784 GGTGTAAACGGGGGAGCTACGGGGGCTACATGACCAACGTTTACTTAACCAAAACTAGA 172843
```

```
QY 1597 CGCTTTAAAGCTGCCGTTTACCAGCGCTCGATATCAAAATTTGGATCAGCTTTCACGGCGTC 1656
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172844 AGGTTCAAGCGCGCGGTACACACAGAGGAGCTGCAGCAACTGGATAAGCTTCTTATGGGAG 172903

QY 1657 AGTGATATCG --- GCTATTCTTTACAGACTGGCAGCTTGAGCATGACATGTTTGGAGAC 1713
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172904 AGCGACATAGGATGGTATTTTCGCCCGGAACTTATCTCCGCGCAAGAACCATTGGAGAGAC 172963

QY 1714 ACAGAAAGCTCTGGGACCGGTCTCTTTTAAATACCGACGAAACGTTGGAGACACCGCTT 1773
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172964 CTGAGAAATACCTTGACTTTAGCCGCTCTTCTCTCTTGAGATAATTAGACACCACTG 173023

QY 1774 TTGATACTGCATGGCAGCGGATGACCGATGCCGATGCCGATCGAGCGGAGCAGCAGCTGTT 1833
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173024 CTTATATACATCCACAGAGGACTTTAGATGCCCTCAGCGGTGATTTCAACTCTTC 173083

QY 1834 ATCGCTCTGAAAAAATGGCAAGAAACCAAGCTTGTCTGCTTTTCCGAAATGTCATCGCAC 1893
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173084 ACCGCTTAAAGCTGAAGGCTGGAGACCAGGCTTCTAGTGTTCAGGTGAGAACCAT 173143

QY 1894 AATTTATCAGCACCGACACCAAGACACGCGATCAAGCGCTTGAATATATACAGTCA 1953
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173144 AACCTCTCCGAAGCGGAACCTCTAAGAGCGGCTGGCAAGGCTTAGGGCGATAGCCTCC 173203

QY 1954 TGGTTTGATCAACATCT 1970
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173204 TGGTTCAAGGAACACCT 173220

RESULT 13
AE010617 9531 bp DNA linear BCT 25-MAR-2002
Fusobacterium nucleatum subsp. nucleatum ATCC 25586, section 159 of
197 of the complete genome.
ACCESSION AE010617 AE009951
VERSION AE010617.1 GI:19714722
KEYWORDS
SOURCE Fusobacterium nucleatum subsp. nucleatum ATCC 25586.
Fusobacterium nucleatum subsp. nucleatum ATCC 25586.
Bacteria; Fusobacteriia; Fusobacterium.
REFERENCE 1 (bases 1 to 9531)
AUTHORS Kapral,V., Anderson,I., Ivanova,N., Reznik,G., Los,T.,
Lykidis,A., Bhattacharyya,A., Bartman,A., Gardner,W., Grechkin,G.,
Zhu,L., Vasileva,O., Chu,L., Kogan,Y., Chaga,O., Goltsman,E.,
Bernal,A., Larsen,N., D'Souza,M., Walunas,T., Pusch,G.,
Haselkorn,R., Fonstein,M., Kyrpides,N. and Overbeek,R.
Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586
J. Bacteriol. 184 (7), 2005-2018 (2002)
JOURNAL MEDLINE
21886394
PUBMED 1189109
REFERENCE 2 (bases 1 to 9531)
AUTHORS Kapral,V., Anderson,I., Ivanova,N., Reznik,G., Los,T.,
Lykidis,A., Bhattacharyya,A., Bartman,A., Gardner,W., Grechkin,G.,
Zhu,L., Chu,L., Kogan,Y., Chaga,O., Goltsman,E., Bernal,A.,
Larsen,N., D'Souza,M., Walunas,T., Pusch,G.D., Haselkorn,R.,
Fonstein,M., Kyrpides,N. and Overbeek,R.
Direct Submission
Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park
Drive, Chicago, IL 60612, USA
FEATURES
source
1. .9531
/organism="Fusobacterium nucleatum subsp. nucleatum ATCC
25586"
/strain="ATCC 25586"
/sub_species="nucleatum"
/db_xref="ATCC:25586"
/db_xref="taxon:190304"
complement(114..2600)
/gene="FN1122"
complement(114..2600)
/gene="FN1122"
```

```
/EC_number="6.2.1.3"
/codon_start=1
/transl_table=11
/product="Long-chain-fatty-acid--CoA ligase"
/protein_id="AAL95318.1"
/db_xref="GI:19714723"
/translation="MQIVTDKNKVALYKDIATYKEFIILNTKKIKQFTKIKEFTNNM
IYVENPELLYSFAIWDSCRATCVDASSTAVELTVYIDNSDVVKIFTTKMQVEKVK
GALISLSKOIEIIVDEINLSEIKIDENSENELVINSPEKEDFALILYTSCTGCKPKG
VMTPTNLANLWSDLVYKMYEEDVTIALLPLHLIPLIGTGWMPLLYSATIVFLED
ISSVALIDAMKVKGTMMGICVPLKWEVHKRIMDTINSKATRFIFKLAKKVNLSFS
KLIFKVSFGFGRHIFKFSVGGSKLNPQITKDFVTGKIKICEGYMTETSIISYTPK
NNTFVSAGKVIKDVVEKTAADNEILVKGKRVKGYKNEPATVEIIDKQGLHTGOL
KPLDNYLITGKMKEMIVLSNGKINPIEIEIKISSMTNLSISIVTEYNSILTAIV
HPLEKVEKVDNVIENLKWVYKQKTPDYKKILDVKILNEDFPKIKGKKEF
MTADMLDKIEKQERPEPFEEYKIKKYLVDIKGDYVDFSHIEDLGMDSLDWVE
FQFLDNLGKEENLSIKPTLLELANIYKGNRQEKIGNLQWKEIVKNDYTKAKLS
SSFIATIFLPSIFPTFRVVKKEKIEKDKPTIFVANQHSGFDFLNFYVPLK
VLKTYFLATVIFKFSIMKFMADSSNVVLVMDNMDIAEVMQILAKLKENKNTAIYP
EGRTDRGKMKKFKFALAKELINIDQPVVIDGAYDLFPAGKKFPPGPKIRVEFLD
KIKVEKLTVIDEIVNAYSIVKNIKIESDTN"
2835..3308
/gene="FN1123"
CDS
2835..3308
/gene="FN1123"
/codon_start=1
/transl_table=11
/product="Thioredoxin-like protein"
/protein_id="AAL95319.1"
/db_xref="GI:19714724"
/translation="MKDGNVTLPNIELVQYGGKHNLEDYKGVVMINFWVWSCDC
KGMKPVVLEYKEYGKNKDLIILGVATPIISKEYPNKDKIDKALLKYIADNRKVFPP
SLPDETGVKVAEIEEYPTFTIDRNGHLKVIYKGAISKEELQNTIESVLNPKK"
complement(3365..4312)
/gene="FN1124"
complement(3365..4312)
/gene="FN1124"
/codon_start=1
/transl_table=11
/product="Outer membrane porin F"
/protein_id="AAL95320.1"
/db_xref="GI:19714725"
/translation="MTKKISSVPGISVTPVKVNSNIVGGDYSMQIKNGDQFTDKD
KTVQDCKAGQYTDENITIRNEDSGQYTKVGTVLQVDTQEGQYLDKRNKISF
QIGADGTVPKDNNTTITIGNNSTYTKGNITIGNNGSGTGNKEKELLENDG
KGAITTLGKTEVEAKPLEKPEKPKLMPVPVPSIEANSLITIDSGILFDVQKY
DVPEAEVLKNIIVLVKADIKAFEDIDGTDSDADEHNQVLSNANAVKNEFTSQ
GMAETIKGYSRPTASNDTPEGKKNRRVEIPII"
complement(4567..5118)
/gene="FN1125"
complement(4567..5118)
/gene="FN1125"
/codon_start=1
/transl_table=11
/product="Lema protein"
/protein_id="AAL95321.1"
/db_xref="GI:19714726"
/translation="MIALGVIIIGIIVILAGIYKKNKVVLDNRVKNSSOIDVQMQ
NPSLVPNIYETVKYAKHEKTEFEGIANAKRYMSATPEEKANQOLSGFLGRLE
AISEAPPELKANSFNENLQQLVEVENKIRFARQFYNDTVTEYNQTIOMPPGSLFAGF
FNTHNAELFKANEMAREEVQVF"
complement(5127..5594)
/gene="FN1126"
complement(5127..5594)
/gene="FN1126"
/EC_number="2.3.1.-"
/codon_start=1
/transl_table=11
/product="Acetyltransferase"
/protein_id="AAL95322.1"
/db_xref="GI:19714727"
/translation="MKSDYDIIIEIKKDNINLIKDLWEKNRIFHONKTSNFSYOVLNLI
FNERMNIIFNSKNIKYIKSAIINKSNIVGYCLSIIOGSSGELCTLFLIDBOHRNGLG
```

```
HLLVDKHLDLKDNKCESIFVNVLVENESTISFYELGFKQNIINMEIPLKKI"
complement(5614..7434)
/gene="FN1127"
complement(5614..7434)
/gene="FN1127"
/codon_start=1
/transl_table=11
/product="Hypothetical membrane-spanning protein"
/protein_id="AAL95323.1"
/db_xref="GI:19714728"
/translation="MKKNILRILFLIISIISFASFSISDLDEAKLQKDGSMVSE
ATYVIDINGIYFDIDAKGYGITSLOVFEDECHYENNVIYSREVDPVNYEVTENDC
VYRIKLYSKNNVRTFKVYLPPEAKIVYNDVLAQLNRKMGVDWQOGISVTKVITEL
PVNSYDNKSVLVFGHPLTGVDKLVNTVYRDDYVPGDFLEAHLIMEPVIFSEFN
VSNVYTKNMKQELLDEARLADNEERDNALREENIOKLSNAKTIFGVEASIVAV
LMYYIHVYFKRKRKSKNDIKYLRDLDDSSPALVGGVATKSYNNNEILATIVDLIRK
LVITLTSKKYIITLTSGTGLSQAOKETIIDIYDFDGRSLDLKSGIFFHKVPMK
TAGKEKSSYIINEMNRGLIYEHGCGATLIFILLSVIFAGGLQATLNTNLFEM
FGIPLGVLDFFSAGTAKYPSKLAETISKWQAFNFLSDYSOLEEAKITSIHLEWQYF
VYAIAGVSDKVVKAYRKALDMGIKIDSGISNITSYPIFNNNFSRNFNLNGMVSKT
NSRASAATASSRRSSSGGGGFSGSSGGSGGSRGGGGG"
complement(7460..9442)
/gene="FN1128"
complement(7460..9442)
/gene="FN1128"
/EC_number="3.4.19.1"
/codon_start=1
/transl_table=11
/product="Acylamino-acid-releasing enzyme"
/protein_id="AAL95324.1"
/db_xref="GI:19714729"
/translation="MENLKDSFLKEYKFLSLNDFDGSNLAFSISBADLETNSYKHF
IYNLDTKKEIKKLTHSGEKNSLNNTTILFSAGDKDIEEKKRIGETWTIIFYALD
IKNGCEAVEYIKLPVNVTEIKIIDENNFILTADFDNNSLNLDLKGPEREKAIOIEE
NKDYVELDEIPFWSNGNGERNKRRLYHFDKLNKLIPLISDEYTNVESVNRKENVI
FIRYIKDKQTLTLAGTIDYKNNKLETIIPDNLIDYDISTANFIEDKICALSDMKFEG
INENHKYILIDSKNINLTYENDTILACTVSDCLGGGSKFVIGNKLYFLSTIADS
VHLSLDTNGKVEVSENGSIDFDIANNIYIVGMRYNLSLOEIKYLENNSIKLSS
FNEKINKKYIKISPEVDFITNGDTKGFVIYIDDKNTYPAIILDIGHGPKTVYGD
VYHEMVQWANNMGYVIFTNPHGSDGYKNKFADIRGYKTIDYEDLNMFTDYLEKYP
TDKSRVGTGSGSYGGVMTNIIIGHTDRFKCAASORSISNWSKFTGTDICYFNADON
QATPINHDKLWHSPLKYADKAKTETLFIHSEEDYRCWLAEGLMETALKYHGEAR
LCMFNGENHELRSRGPKRIRRLTEITNWFEXYK"
BASE COUNT 3195 a 1514 c 965 g 3857 t
ORIGIN
Query Match 8.1%; Score 159.6; DB 1; Length 9531;
Best Local Similarity 52.5%; Pred. No. 2e-39;
Matches 423; Conservative 0; Mismatches 374; Indels 9; Gaps 3;
QY 1165 AATGACAGTTTGTGTCAGGAGCATACGATATCAATACCTGAGAGATTCATATGCTACA 1224
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8263 AATGAAAAATAAATAAAAAATATAGATATCAAGCCAGATTTTCGATTTTATTAC- 8205
QY 1225 GAAGACGGCGTGATGCTGAACGGCTGGCTGATGAGCGCTGCCAAATGGAAGTGAGACA 1284
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8204 -AAATGGAGATACACAAAAAGGTTTGTAAATTTATCCTATTGACTATGATAAAAAATAA 8147
QY 1285 ACATATCCACTTATCTTAACTACACACGGCGGCGCATATGATGTCGACGACATACATAT 1344
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8146 ACTTATCTCGAATACTTGATATACATGATGTTGGACCTTAAACACATTTTATGAGATGTTTAT 8087
QY 1345 TTTTCATGAGTTTTCAGTGTGCGCGGAAAGGATACGCGGTCTGTTTATATCAATCCGAGA 1404
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8086 TATCATGAATCAAGTTTGGCAATATATGGTTACTTTCGTAATTTTACTAATCCACAT 8027
QY 1405 GGAAGCCACGGCTACGGCGAGGAATTTGTGAATGTCGGGTGACAGAGATTTATGGGGGAAG 1464
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8026 GGTAGTGATGATATGGAATAAATTTGCAGAT---ATAGAGAGAAAATACGGGAACAATT 7970
QY 1465 GATTATGACGATGTGATGACGCTGTGGATGAGGCTATCAACACGAGATCCGCATATTGAT 1524
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7969 GACTATGAAGATTTAATGAATTTCACTGATGATTGTTTTAGAAAAATATCC---TATTGAT 7913
```

```
QY 1525 CCTAAGCGCTCGGTGTACCGGGCGGAAGCTACGAGGTTTTATGACCAACTGGATCGTC 1584
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7912 AAATCAAGATTTGGGGTACAGGTGGTCTTATGTGGCTATATGACTAACTGGATAATA 7853
QY 1585 GGGCAGAGAAACCGCTTTAAAGCTCCGTTACCCAGCGCTCGATATCAAAATTTGGATCAGC 1644
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7852 GGACATACAGATAGATTAAAGTGGCGAGCCCTCTCAAGAAGTATATCTAACTGGATTCA 7793
QY 1645 TTTCACGGGTCAGTGATATCGGCTATTTCTTTACAGACTGGCAGCTTTGAGCATGACATG 1704
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7792 AAATTTGGTACACAGATATTGGCTATTACTTTAATGCTGACCAAAACCAAGCTACTCCT 7733
QY 1705 TTTCAGGACACAGAAAGCTCTGGGACCGGCTCTCTTTAAATACGCGCAACAGTGGAG 1764
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7732 TGGATAATCATGATAAATATGTTGGGCACTTCCACCTTAATATGCGAGATAAAGCTAAA 7673
QY 1765 ACACGCTTTTATGACTCGATCGGAGCGGATGACCGGATGCGGCTGAGCAGCGCGAG 1824
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7672 ACACCAACTTTATTTATTCATTCAGAAGAAGATTATAGATGTTGTTAGCAGAGGATTA 7613
QY 1825 CAGCTGTTTATCGCTCTGAAAATAATGGCAAGGAACCAAGCTTGTCCGTTTCCGAAT 1884
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7612 CAAATGTTTACTGCTTTAAATATATCATGGAATAGAGCTAGACTTTGTATGTTAGAGGA 7553
QY 1885 GCATCGCAACAATTTATCACGCGGACACCCAGACAGCGGATGCGGCTGCAATTTAT 1944
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7552 GAAATCATGATTTATCATCAAGAAGTGGAAACCTTAACATAGATTTAGAAGATTGACAGAA 7493
QY 1945 ATCAGCTCATGTTTATGATCAACATCT 1970
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
7492 ATTACAAATTTGTTTGAATAATTT 7467

RESULT 14
AE009798 Pyrobaculum aerophilum strain IM2 DNA linear BCT 16-JAN-2002
LOCUS Pyrobaculum aerophilum
DEFINITION genome.
ACCESSION AE009798
VERSION AE009798.1 GI:18159783
KEYWORDS Pyrobaculum aerophilum.
SOURCE Pyrobaculum aerophilum.
ORGANISM Pyrobaculum aerophilum.
REFERENCE 1 (bases 1 to 11233)
AUTHORS Fitz-Gibbon,S.T., Ladner,H., Kim,U.J., Stetter,K.O., Simon,M.I. and Miller,J.H.
TITLE Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 984-989 (2002)
PUBMED 11792869
REFERENCE 2 (bases 1 to 11233)
AUTHORS Fitz-Gibbon,S.T., Ladner,H., Kim,U.-J., Stetter,K.O., Simon,M.I. and Miller,J.H.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2001) Microbiology and Molecular Genetics, University of California, Los Angeles, 405 Hilgard Ave, Los Angeles, CA 90095-1489, USA
FEATURES
source
1. .11233
/organism="Pyrobaculum aerophilum"
/strain="IM2"
/db_xref="taxon:13773"
258..695
/genes="PAE0993"
258..695
/genes="PAE0993"
/notes="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="conserved within P. aerophilum"
gene
CDS
```

```
/protein_id="AAL63184.1"
/db_xref="GI:18159784"
/translation="MIFDVLGHVVFVKARFVGDRGDKIADALIDFGATYTVIIPRELA
BEIGAPRTGIYAEVKTAGVVKLELTHILLEIEGTALAEAVLSLESLPIGLVRIE
GLGRVDPTTGKLESLRNFGLAISITILSRISFRIGRIYVCL"
complement(676..999)
/genes="PAE0995"
complement(676..999)
/genes="PAE0995"
/notes="Hypothetical"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL63185.1"
/db_xref="GI:18159785"
/translation="MDIDKINECKFKYVRHNI FVSLDMDTKRDELTKRDELTKALCDCE
RPRINAEVPHKTAEGAFHPRLEALVARGWKLEALAGPIPISCLLYPYRLREITIL
IDRR1"
complement(1021..1557)
/genes="PAE0996"
complement(1021..1557)
/genes="PAE0996"
/notes="Hypothetical"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL63186.1"
/db_xref="GI:18159786"
/translation="MKISKSTLHGDEKSRMDRTDLRLVKKLDKDFKKNKIDINRIY
BEIEFYRGERLGDGIEFAKRIALAEHHTKDKLKIOTELRRAYKADMKREKAYTT
VAAGOMLIPNLIVSHPAEDWFERLULLAFAGIGAVNYEIKDGETILFEHAGRRKIED
VKIKNSQLVTSQVLSL"
complement(1729..3183)
/genes="PAE0998"
complement(1729..3183)
/genes="PAE0998"
/notes="Hypothetical; Conserved"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL63187.1"
/db_xref="GI:18159787"
/translation="WRNIPINKINDYVWEIPGVPCKVPVRIYADSVLLEKMSDM
TLEQINVCGLPGIYRWSIVLDAHQGYFGPGVAAIDAEGVISPGGIDVINCQV
RVLRTNLTEEDVRPKLELVDITIFLVPVPGGTLRLSLSEFVRLAEVGEVAVQK
GYWAEDEMYIEERGSKWLDPSKVKSEKAKGRDQGLTSGNHFLETVQVVDKIYDE
KIAKLFGEREQGVVMIHTSGRGFHOVATDYLLIMERKMROWGLNLPDRLEAAAPL
KDKVAEDYIKAMASAAFPATNRHTIIMHWREAFKVFSGTEKVGLEVVDVAHNIK
LEEHVDEKGTVRKVWVHRKATRAFPGRSEIPAKYREVGPVLLIPSGMGTASHILY
GTHDAMRLTGTAPHGAGRVLSREAAIRMPYHKVOEEMAKRGIIVRSAEETVISEEA
PWATKDVDRVVEAAHQVGAQKVRQRPVGK"
3425..5626
/genes="PAE0999"
3425..5626
/genes="PAE0999"
/notes="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="P. aerophilum family 59 protein"
/protein_id="AAL63188.1"
/db_xref="GI:18159788"
/translation="MIRVLIATVLLAAGQIPGKPLICLNASAGLEGSPML
AGFVVGKGGQVQLYVNGTVINVGWDANSKPKQIGNITVLPPIAKGLICFGPDG
VPARVARGLYLEARGRRYGTIINASDDPAAPVAPVLEYPDRVYIRGRVVAKRE
VGERGFAGLGRVEVMQPSQSLFRDFSGYIIASFKNLYTDNYLNLYGSEACT
KGRF ILPNGTGWFTLAFTPATRADYGGASYLYLTIRASVYDADTDLPLANTIGNIY
VPNPRPHLVTIDLSQWTLTKRIDLTICKPSGVYVGVYIDAVYAAVPAKSYAE
PFALESPPRTLYKPTGSSQLFNGTYLAPGLPFPYGGALGSARSLVITKCSFRAP
SSVNIYMGPTIIGTVSRIQDGCYIVVTFADRGDDAAVRSSLYLGGTLHAYYIGP
FNNDVYNTGLRIDQLRIEGLYRPEMNARIFKDYATNMGYMTLSFVYTYVTPNV
PIIATYGSKVDIKFSGLSYTRPLIIAVQKYGDMWHVCGSISFVVRAYSNGQPTLDP
DGTAGVYQCGNSDLIWLVDVFIISNLDEAKNAVSQLVKVYSWYSFTIGFVQAA
FVGVDVRVKTODGGVSYTISIGSLAPDAIVIDGATFSLSPSDAIEPIERIEIQTSCAG
```

```
gene
5619. 6149
/ gene="PAE1000"
CDS
5619. 6149
/ note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="P. aerophilum family 3 protein"
/protein_id="AAL63189.1"
/db_xref="GI:18159789"
/translation="MARLFKYPFRFMGLALGVMSFFHLFLFGLREGLLAAVVSFYS
AYGVYNNRWLVWGVAVAGVFISGRIVVVFVFLYLAEMLAVPLFLTRGVRSRVKS
LFEFLVGLHSLGSLVPLLLSLPLWRSVLPFFEFIAVGAIPTPVDWPFVGVYVWERY
HSIKGLPSGFTREGE"
6290. 6733
/ gene="PAE1001"
CDS
6290. 6733
/ note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="conserved within P. aerophilum"
/protein_id="AAL63190.1"
/db_xref="GI:18159790"
/translation="MRGAVVSIQVAVLLVFLIAVLAAGLYLTYFYSSLOVYIALT
OAYTPPSGAENVKICFSGVSGGLTITGVLDGRPAASEFKVYVRGAEGAEVRAGDVG
YIRADFGVVVRPGWMPTRVTVRQGSFLTPSPVQAEGRCPE"
6760. 8643
/ gene="PAE1002"
CDS
6760. 8643
/ note="Protein fate; Protein modification and repair"
/codon_start=1
/transl_table=11
/product="acylamino-acid-releasing enzyme, conjectural"
/protein_id="AAL63191.1"
/db_xref="GI:18159791"
/translation="MSLSPEDLAKLVLSNPDAAGKGVFFVTTRISLENDREYSIIWA
CEGVCKVPLPGPDTAARAPDGSKIAPLSRRGFGDKGVGLWAVSWGEPROLAK
FLGVLDSPSGGALAVAYEGSPGVKVERLPLWINDGFAYNYSLLYLVDSY
SGVKEITDGVVRKAAWADPRRIATYAVDLKQLQDVLVDLTKIGEQMTAASG
FSSVSDIAMSPPDHIAFTHLRPLGSLSHSRVLEEGEGPRVCVGFVNVGNSVN
SDVPSFARALYAGDGLLFOATVGVTVGFRAALSDGVDFVLPFGVDEFLVGG
AIVTYMEANPKELY IWDGEPRLTRFNDVFLERWKLRRPSRLVLRASDGAEEVG
VLLPVGCPKHWLY IHGGPKTAYGEGMFEFHLASRGYAVFNSPRGSDGYSEFA
DIRKYGERRDQDLMEVVDVYVLPDNRCAVAGSGYGFMTWIIAHVDRFAAV
TORSCDVMISGTTIDIGYFVEDQLCCTPWRNRELCEKSPSLYLANRKYTPILLHS
LEDYRTWVQGVLFPTALRLHGVETRLVLFPEESHELTRKKGPRHRVENFKELLNLD
KYV"
8636. 9889
/ gene="PAE1003"
CDS
8636. 9889
/ note="Amino acid biosynthesis; Aspartate family"
/codon_start=1
/transl_table=11
/product="asparaginase (asnA)"
/protein_id="AAL63192.1"
/db_xref="GI:18159792"
/translation="MKKRVVLENGDVFEGVWMLPPTQFSDSDIVVLKKNYVGFK
KGRKEIVELGEVQTAPLSAKPMPKIEGKFWLLATGGTILSRVDYVTGGVYPTLSVD
YLFEVLGLEAPIEAEVYAKFSEDMTPALMGVIAERVAEAFKKGARGVVLHGDTM
QYTAALAFAPKSPGPIALVGAQRSDSPDVAVNLKAAIAVTAARAPFAESVVMH
KTSQGVTVHRTVRKMTSRRTFOSINTPIAEYPEKELLQVLVDYKERGGL
DYAKFEAAVALYKFPYCPMHPLLEALLVGMKGVVIEGTGFGHVGEGVLPVAKKLIID
AGVIVMTSQILYGRVNLVYVRGRELISMGVIPLEMLPETAYAKMSWALANFREER
VPRVLTTPPIAYEMSPRSDPLVFGGL"
9886. 10416
/ gene="PAE1004"
CDS
9886. 10416
/ note="Hypothetical; Conserved"
/ codon_start=1
/ transl_table=11
/ product="conserved within P. aerophilum"
/ protein_id="AAL63193.1"
/ db_xref="GI:18159793"
/ translation="MNDVIAVADTLCKTPRVGLQVGVSEAESVCAHVLLTLLAGEI
AAHLNAGVEVDIAKVLAVAADHDLAESVLGHPGREVDRKIRWEAEVEEFKRFPHL
AELFRWRYETNLVGRIVSFADLALIRACRYKKLGVDLDDLIKLYRLKLYDDLLA
HILDIYVARYCOGVLE"
BASE COUNT 2438 a 2847 c 3137 g 2811 t
ORIGIN
Query Match 8.0%; Score 158.2; DB 1; Length 11233;
Best Local Similarity 54.4%; Pred. No. 6e-39;
Matches 363; Conservative 0; Mismatches 298; Indels 6; Gaps 2;
QY 1304 ACATACACGGCGTCGCGCATATGATGTACGGACATACATATTTTCATGAGTTTCAGGTGC 1363
|||||
DB 7979 ACATACACGGCGGGGGCCCAAAACAGCCCTACGGCGAGGGGTTTATGTTGAATCCACTTGC 8038
|||||
QY 1364 TGGCGCGGAAGGATACGGCGTCTTATATCAATCCGAGAGGAGCCACGGCTACGGGC 1423
|||||
DB 8039 TCGCGTCGGCGGATACGCTGTGTCTCAAAACCCGAGGGGAGCGGCTATACGC 8098
|||||
QY 1424 AGGAATTTGGAATCGGTCAGAGGAGATATATGGGGGAAGGATATGACGATGTGTC 1483
|||||
DB 8099 AGGAGTTTGGCGAT---ATTAGTGTTAAATACGGCGAGAGGGACTTTCAAGACTTAATGC 8155
|||||
QY 1484 AGGCTGTGATGAGGCTATCAACGAGATCCGCAATTATGATCTTAAGCGGCTCGGTGTCA 1543
|||||
DB 8156 AGTTGTGGACTACGTGCTTAAGAGCTATCGGC---TTGACCCCAATCGAGCGCTGTGG 8212
|||||
QY 1544 CGGCGGAAAGCTACGGAGGTTTATGACCAACTGGATCGTCGGCGAGACAGAACCGCTTTA 1603
|||||
DB 8213 CCGCGGGTGTACGGCGGCTTTATGACGAATTTGATAATTCGCCACGCTGGACAGGTTTA 8272
|||||
QY 1604 AGCTCGCTTACCGAGCGCTCGATATCAATTCGATGAGTTCACGGCGCTCAGTGATA 1663
|||||
DB 8273 GAGCCGCGGTGACGCGAGCGCTCAATATGCGACTGGGTGTCTATGTACGGCACGCGGATA 8332
|||||
QY 1664 TCGGCTATTCTTTACAGCTGGCAGCTTGAGCATGATCATCTTTTGAGGACACAGAAAAAGC 1723
|||||
DB 8333 TAGGTTGGTATTGTCGAGGATCAGCTCTGTTGCACTCCCTGGAGGAACAGGAGCTCT 8392
|||||
QY 1724 TCTGGGACCGGCTCTCTTTTAAATACGAGCAACGAGGAGGACACCGCTTTTGATCTGC 1783
|||||
DB 8393 CGCTTGAGAAGAGCCCGCTTTATTTAGCCATAGGGTAAAGACGCCGACATTAATAATCC 8452
|||||
QY 1784 ATGGCGAGCGGGATGACCGATGCCCGATCGAGAGCGGAGCAGCTGTTTATCGCTCTGA 1843
|||||
DB 8453 ACTCTCTAGAGGACTACAGGACGTGGGTAGACCGGGGGTGTGTTCTTACGCGCATTA 8512
|||||
QY 1844 AAAAAATGGGCAAGGAAACCAAGCTTGTCCGCTTTTCCGAATGCAATCGCACAAATTTATCAC 1903
|||||
DB 8513 GGCTTACGGGTTGAGACGCGGCTGTTCTTATTTCCCGAGGAGAGCCAGCAGCTGACAC 8572
|||||
QY 1904 GCACCGGACACCCAAAGACAGCGGATCAAGCCCTGAATATATATGCTCATGTTGATC 1963
|||||
DB 8573 GCAAGGGAAGCCCGCCACCGGGTGGAAAACTTTAAGGAGATCTTAAGCTGGCTGGATA 8632
|||||
QY 1964 AACATCT 1970
DB 8633 AGTATGT 8639
```

```
RESULT 15
CNSPAX05/c 304290 bp DNA linear BCT 07-AUG-2000
LOCUS Pyrococcus abyssi complete genome; segment 5/6.
DEFINITION AJ248287 AL096836
ACCESSION AJ248287.1 GI:5458657
VERSION complete genome.
KEYWORDS Pyrococcus abyssi.
SOURCE
```

ORGANISM

Pyrococcus abyssi
Archaea: Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
1 (bases 1 to 304290)
Heilig, R.
Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution
Unpublished
2 (bases 1 to 304290)
Genoscope.
Direct Submission
Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Location/Qualifiers

FEATURES
source

1..304290
/organism="Pyrococcus abyssi"
/strain="Orsay"
/db_xref="taxon:29292"
complement(77..2398)
/note="PAB1560"
/codon_start=1
/transl_table=11
/product="hydrogenase expression/formation regulatory
protein (hypf)"
/protein_id="CAB50145.1"
/db_xref="GI:5458658"
/db_xref="SPREMBL:Q9UZB4"
/translation="MKAYRIHVQIOVQVGRFPVPIAHANLGRVYRLGDAGVEI
VVGREDIEGLRDYRKPPIAKIEKVEEPIQGFDFRYEIKSSTKKGGDSI
IPDIAEDCELEDFNTKRYMYPVICTNGCPFTIIEDLPYDRENTAMRFPYMC
DFCKSEYEDLPNRYHAEPVACVPGSPYLYTSDGKEITGDLRKAALDKGYIITA
IKGIGGTACADATREDVAELAKRLPRCPKPAIMAKDLTVKGFAYISREEEELT
SYRPIVALKRPEPLPENLAPCLHTIGWMLPYAGTHYILFHWKTPVYVMTSANEP
GMPKIDNEAFKLDVADYLLHNRRINRADDVVRVDDGRRVIRSRGFPVLPW
IEIPFTEKVAELMNAFGKANGKVPSPQYIGNTSKIEVLEEMAEIKHFFKILR
INALDVLADHPNNTKLMAIEAEFAEFNQVQHYYAHVASMAEHLNLEEVGIA
LDGVTGDTKWTGTEIYLSYEDIERLAHLEYPLPGDLASYPLURALIGIYLSNH
EIEVENTIRKFCPNALSKYGETELKVTILROLSSGINVAYASSTGRVIDADSVLLN
VSYRRYEGEPAMKLESFASQGNLDKLTIPVGEETIRVSELFEELDLMGKASPRDI
AYSILHALAFAEVSVEKAKGKATVVLGGVGVNELIVTKRIVEGNLGRKLFLLT
YEVPRDNGINVGQAFGLGLYLEGILNREDLST"
complement(2490..3593)
/note="PAB1559"
/codon_start=1
/transl_table=11
/product="hydrogenase expression/formation protein (hypD)"
/protein_id="CAB50146.1"
/db_xref="GI:5458659"
/db_xref="SPREMBL:Q9UZB3"
/translation="MLEKPRDKIEIAOKIVRKIHKAKGFDLRFMHVCGTHEDVTTRA
GIRSLPENVKIMSGPCPVCITPVEDIVRMMEIMKAKEGEDLILTTFGDMYRPT
PMGSFADLSEGDVRYVYSIYSYKIAQENRDKLVVHFSFGFETITAPAGMLETVI
NEGDLNFKYSVRLRPVPAEALLKAGTVPHGLIDPHVSTIIGVRGWAYLTETKNIP
QVVAPEPDLVLGLILIRMKVEGAKEIINEYSRVVKNEGNKAQELINKFPEVDA
KRWALGIIPDSGLELKEWKDLERTLYNPDPVKLPDLKGLCGLCGAILRLGALPPQCP
HFCKTTPRHPIGPMVSVGTCISIFYKGALE"
complement(3594..3821)
/note="PAB7315"
/codon_start=1
/transl_table=11
/product="hydrogenase expression/formation protein (hypC)"
/protein_id="CAB50147.1"
/db_xref="GI:5458660"
/db_xref="SPREMBL:Q9UZB2"
/translation="MCLAIPIGRVVRVENVGIVDFGVKREVRDLDSLVKVGVDYIV
HTGFALKEUDEERAKETLEAWDEVFSILGSE"
3949..4386
/note="PAB0821"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50148.1"

CDS

/db_xref="GI:5458661"
/db_xref="SPREMBL:Q9UZB1"
/translation="MODSDSEIOELKRNLOVAFEVLAEIAQSVIRLNLVAYEGGI
GIDVAIPENDPIAREIVKILFDLRRANVSQITRELKRGKASRNTVRAKLNLKLN
IGVVVEVPGERGKVTLSKDVVKRWLDLIGIPKLEHKNDYLR"
4392..4856
/note="PAB0822"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50149.1"
/db_xref="GI:5458662"
/db_xref="SPREMBL:Q9UZB0"
/translation="MEKEKLEKVEELIEIKSAKTEEEVEVLRKRVAKVKLIAYE
GDKDLTKAMLDKLPVMDLLEPLQKLADLYSPKMOALGKSVQAFYKDLVEAGM
DKETATLTKEFMKNINVMRSLEGLTGMLOKGGFQVYVHGKKEKRETKS"
complement(5039..5668)
/note="PAB1558"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50150.1"
/db_xref="GI:5458663"
/db_xref="SPREMBL:Q9UZA9"
/translation="MLKEIFNGIRFGKEITALEYDSNTPTVYVHLLSWSEKGYKIV
IFDVFSLIVYRRKAEELGFDTSCKAKVYKGVGRVNEGVNVMKLTVTYEGPPEKVL
FEATEOVYEDKTIIVIPGTEKVLSTLYPPREQLSFTNFIARVGDKRKAFHINTDLV
KAVAPQILPMLEESFTVMRLKYGKVEKLIWVKSINPKLDGLEVLAKS"
complement(5709..6227)
/note="PAB1557"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50151.1"
/db_xref="GI:5458664"
/db_xref="SPREMBL:Q9UZA8"
/translation="MEGRVLLYPMRTYLVVSGHDEANVMAADVTVLSTDFPFVGVGA
IAPRTTHKLKKNYEPVISPSEMLRDWVAGTKGPKSKDKMGITLVVPSRKVKVP
SIRKALANICEVEYTDYDGHDTLVFGVKVLAITYKEAEFEGKPNLKFNLARHWSWE
FVTFQDIYRAE"
complement(6269..6913)
/note="PAB1556"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50152.1"
/db_xref="GI:5458665"
/db_xref="SPREMBL:Q9UZA7"
/translation="MSVEFSTMILLIMVLAAVSTLQYKGRKLNLLMQHLYLSIEAV
VKPKDXYVMVGXYGFRAYKKLEDKDLKFEYTLTLPLPROSILYFPLSLINRHKLL
YVVRPTDKREVHFKQGYFRMKVIRKIEILLKGEIEINGVKFYKYEKSRDVEA
VKPEISGFSKVKVKNHVALPKNTVIYVFFKPEPESEDVVRNIRIYKVRMANR"
complement(7184..8176)
/note="PAB1555"
/codon_start=1
/transl_table=11
/product="ANION TRANSPORTING ATPASE"
/protein_id="CAB50153.1"
/db_xref="GI:5458666"
/db_xref="SPREMBL:Q9UZA6"
/translation="MREFLTPKEGFRVIFVIGKGVGKTTSSASISVALAKAGYRTLI
VSLDPAHNLGDVMEKLSDKPREILDNLYASELMEGMIKGYLEHLEKLNKNMYRTLI
VINLEKFEVLRVSPGLEIATLEAIRLVLKGDWDVIFDTPPTGTLTLVLALPRI
SLWVTEKLDIRKKILERRAITKIQGEOKFVIDGEEKIPLTREETEADVRELKAYRD
EIKFVEDVLTNPNKTSIAYVNMPEFLPYETKRAYESLKKFKIPFMVVMNKVFLKLR
EIPERVKLEAQEKVSEVKRPFPGVEIPEIPQEEPRGIEWLSKVGMETVGR"
complement(8489..10267)
/note="PAB1554"
/codon_start=1
/transl_table=11
/product="CARBON STARVATION PROTEIN A."
/protein_id="CAB50154.1"
/db_xref="GI:5458667"

CDS

CDS

CDS

CDS

/db_xref="SPTREMBL:Q9UZA5"
/translation="MNSTVILAAIIIVGLYFTYGGKLVVVRADPNRPTPAHRLY
DGVYVPAHPLVLYGHFASITAGSRPIVGPVAVAWGWLPSLIWVFNFGVFIAGVHDY
LALMSYRYDCKSIQWITAGLSKRTSMAFELYVWFALILVIAFAAVIAGVFKTPG
AATASILLIALLVLLXKQINFRVGTIGLIIIAIYIGFKPPLVASAKAWYI
FLGAYIVASSLPVWVLLQPDVNLVYILWFLGLGLAFIIGFKVGTETAPAYTT
FSRVVGGKSPWPPTIPLIACGALSQFSLVSGTSSKOLDNELHGLLQVYGMET
EGSLTIVISSIAYVGLQIFHNAGIGITAETWAAQAPLVAAKSITETLQKLGAAL
ENFKNGTVIKGVIFASSYAGLHDAFDIHLGTLFASLWVSAPAFITLDTATLGR
FAWQEIITWVKGPILKNKWASPIVFFGIHALMKWKLWVWPAFSGMNLASIAM
MTASLWMLQRPQSPKWAVALIPALFLWVTVTSALIFLWVGVKSSIVWSLITIVGL
VLNELLVIDATNAWKPAEEYQAAAA"
11287..13068
/note="PAB0824"
/codon_start=1
/transl_table=11
/product="RNASE L INHIBITOR"
/protein_id="CAB50155.1"
/db_xref="GI:5458668"
/db_xref="SPTREMBL:Q9UZA4"
/translation="MYRKRIVADIDYKCNPDKCGHFLCERVCVPMRGGEALIIDEE
NYKPIQEAACSTGGCIVHKCPFNAISIVNPEQDEDCVHRYGYNALVLRPLIVKD
GMVGIYVPGTGTATVAILAGOLIPNLCEEDNSDNIIRAFRGNELONYFERLUNG
EIRVAPQYVLLPKAVKVRRELLKQVKEFVEVYKLELENDRELHQLSGG
ELQVATAALLRAHFVDEPSSYLDIQRKLVARVIRLANEGKAVLVVVEHLAV
LDYLSDVHVYVGPVYVIGFSPKPTRNGINEFLOGYKIDENVRFRYEIRFTKLSE
RDYVERETLVYVPLVDYQFSKLEVEPEIRKGEVIGIVGPNIGKTFVKMLAGVE
EPTGRKVDLTVAKQYIKAEYEGTVYELLSKLNSFYKTELLKPLGLIDL
YDRNEDLSGGELOVAIAATLLRDADIIYLDPEPSAYLDVFORLAVSRAIRHLMKNE
KTALVDEHDLIMDYSDRLVIFEGEPGRHGRALPPMGMRGMMNRFASVGTIFRRDP
DSGRPRANKEGSKVDREOKARGEYVA"
13130..13513
/note="PAB0825"
/codon_start=1
/transl_table=11
/product="TRANSLATION INITIATION INHIBITOR, PUTATIVE."
/protein_id="CAB50156.1"
/db_xref="GI:5458669"
/db_xref="SPTREMBL:Q9UZA3"
/translation="MAKEVIEITEKAPKPIGYSQAIKVGNEIFVAGQIPIDPETGEIV
KGDYKEQTKRVNIENIKALEEAGSLNDVVKVTVYVYLKLDNDFAKMNEVSEYFGESKP
ARVAVEVSRLPKDVLIEMEAIAYPE"
13623..14480
/note="PAB0826"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50157.1"

CDS

CDS

CDS

Query Match 8.0%; Score 158; DB 1; Length 304290;
Best Local Similarity 52.7%; Pred. No. 1.9e-38;
Matches 391; Conservative 0; Mismatches 345; Indels 6; Gaps 2;
QY 1230 CGCGGTGATGTAACGGCTGCTGATGAGCGCTGCACAAATGGAAGGTGAGACACATA 1289
DB 209349 CGAGTTGGAGATAGACGGCTGTGTACATAAAGCCGAGCTCAAGAGGGGCGAGAAGC --- 209293
QY 1290 TCCACTTATCTTAACATACACGCGGTCGCGATATGATGATGACGGACATACATATTTCA 1349
DB 209292 TCCGGTAAATAGTCTTCGTCACGGTGGGCCCAAGGAATGTACGGTTACTACTTCAAGTA 209233
QY 1350 TGAGTTTCAGGTGCTGGCGGCGAAAGGATACGCGGTGCTTTTATCAATCCGAGAGGAAG 1409
DB 209232 CGAGATGACGCTTATGGCGAGCAAGGGCTACTACATAGTCTATGTCAATCCCAAGGGGAAG 209173
QY 1410 CCACGGCTACGGCGAGGATTTGTGATTCGGTTCAGAGGAGATTATGGGGGAAGGATTA 1469
DB 209172 CAACGGTTACACGGAGGACTTTTGACATTAGTTCCTCGAGAGGACTTGGCGCTAGGAGCTT 209113
QY 1470 TCACGATGTGATGAGGCTGTGGATGAGGCTATCAACACGAGATCCGCATATTTGATCCTAA 1529
DB 209112 CCAGGATATCTGAATGGAATAGAGGAGTTCTTCAAGCTTGAACCGCAAGCGGACAGGGA 209053

QY 1530 GCGGCTCGGTGTACGCGGGGGAAGCTACGAGAGTGTATGACCAACTGGATCGTCGGGCA 1589
DB 209052 GAGGGTTGGAATAACGGGAATAAGCTACGAGGCTTTCATGACGAAGTGGGCCCTAACGCA 208993
QY 1590 GACGAACCGCTTTAAAGCTGCCCTTACCCAGCGCTCGATATCAAAATTTGGATCAGCTTTCA 1649
DB 208992 GAGCGACCTATTTCAGGCGCGGAATCAGCGAGAAGGGAATAAGCTACTGGCTAACGAGTTA 208933
QY 1650 CGGCGTCAGTGATATCGGCTATTCTTTACAGACTGCGAGCTTCGAGCATGACATGTTTGA 1709
DB 208932 CGCGTTCTCCGACATAGGATTGGTT ---CGACAAGAGTATTAGTGAGAACCCGCT 208876
QY 1710 GGACACAGAAAAGCTCTGGGACCGGTCTCTTTTAAATACGACGACCAACGTTGGAGACACC 1769
DB 208875 AGNAAACGAGAACTTCAGGAAGCTAAGTCCGCTGTTCTACGTAAAGACGTGAAGGCTCC 208816
QY 1770 GCTTTTGATACTGCATGGCGGAGGATGACCGATGCCCGATCCGATCCAGCAGCGGAGCAGCT 1829
DB 208815 CCTGTTGCTAATCCACAGCCTTGAAGATTATCGCTGTCCATTTGGATCAGAGCTTGATGTT 208756
QY 1830 GTTTATCGCTCTCAAAAAAATGGCAAGGAAACCAAGCTTTCGCTTTTCCGAATGCATC 1889
DB 208755 CTACACAGTCTCGAAGATTTAGCAAGAAGTTTACATACGATATTTCAAGAGGGGAGC 208696
QY 1890 GCACAATTTATACGACCGGACACCCAGCAGCGGATCAAGCGCCTGAATTATATCAG 1949
DB 208695 TCATGGACATAGCATAAAGGGGAAGCCCAAGACACAGGATGAAGAGGTACAAGCTCTTCAT 208636
QY 1950 CTCATGTTTGTATCAACATCTC 1971
DB 208635 GGAGTTCTTCGAGAGGAAGCTC 208614

Search completed: July 4, 2003, 13:12:17

Job time : 5004 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 09:58:29 ; Search time 452 Seconds

(without alignments)
9820.108 Million cell updates/sec

Title: US-09-462-845-1

Perfect score: 1971

Sequence: 1 atgaaagactgataaccgc.....catggttgatcaacatctc 1971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1971	100.0	1971	AA07301	Bacillus subtilis
2	807.8	41.0	1958	ABK72933	Bacillus lichenifo
3	238.2	12.1	265118	AAH41227	Pyrococcus abyssi
4	227.2	11.5	1896	ABV33582	Pyrococcus horikos
5	195.6	9.9	1032	AAQ86198	Lactobacillus rham
6	192.2	9.8	477	ABK77575	Bacillus clausii g
7	158	8.0	349980	AAH41226	Pyrococcus abyssi
8	149.8	7.6	1869	AAV12887	Thermococcus amida
9	82.4	4.2	585	ABK77547	Bacillus clausii g

10	49.8	2.5	659	21	AAF13721	Aspergillus oryzae
11	46.4	2.4	1797	24	ABK73047	Bacillus lichenifo
12	39	2.0	6741	21	AAA10595	Gene encoding a su
13	38.6	2.0	4590	22	AAH24065	Yeast AOD9604-asso
14	37.4	1.9	2196	12	AAQ10958	Gene encoding acyl
15	36.6	1.9	788	24	ABO29152	Oligonucleotide fo
16	36.6	1.9	788	24	ABO29153	Oligonucleotide fo
17	36.4	1.8	2365589	24	ABA90521	Genomic sequence o
18	36.2	1.8	9271	19	AAV23080	Fragment HG2141 o
19	36	1.8	568	24	ABK75665	Bacillus lichenifo
20	35.8	1.8	1395	18	AAV7586	Helicobacter pylor
21	35.4	1.8	716	22	AAH08175	Human cDNA clone (
22	35.2	1.8	8846	22	ABH15292	Human nervous syst
23	34.8	1.8	2196	13	AAO20001	Encodes Human acyl
24	34.6	1.8	45613	22	AAF28535	Genomic fragment #
25	34.2	1.7	729	20	AAZ28861	N.meningitidis nla
26	34.2	1.7	10732	21	AAA10594	Gene encoding a su
27	34.2	1.7	15164	24	ABK35500	Human endometrial
28	33.6	1.7	420	16	AAQ88233	Bovine arGF synthe
29	33.6	1.7	446	21	AAH30728	Human colon cancer
30	33.6	1.7	2301	23	ABL24372	Drosophila melanog
31	33.6	1.7	5828	22	AAF89816	cDNA encoding huma
32	33.6	1.7	5828	24	ABK84150	Human cDNA differe
33	33.6	1.7	5828	24	ABN96524	Gene #3022 used to
34	33.6	1.7	6151	16	AAO75977	Insulin receptor s
35	33.6	1.7	6522	24	ABL33803	Human immune syste
36	33.6	1.7	20300	24	ABK47337	Genomic nucleotide
37	33.4	1.7	744	21	AAZ53852	Neisseria meningit
38	33.4	1.7	957	20	AAZ06344	Cephalosporin acet
39	33.4	1.7	2176	23	AAH89363	DNA encoding novel
40	33.4	1.7	2277	19	AAV05370	Human telomerase p
41	33.4	1.7	3549	22	AAH45086	cDNA encoding nove
42	33.4	1.7	4590	22	AAH24065	Yeast AOD9604-asso
43	33.4	1.7	4917	23	AAH89366	DNA encoding novel
44	33.2	1.7	1406	23	ABL06013	Drosophila melanog
45	33.2	1.7	3480	23	ABL06012	Drosophila melanog

ALIGNMENTS

RESULT 1
AA07301
ID AA07301 standard; DNA; 1971 BP.
AC AA07301;
XX
XX 21-MAY-1999 (first entry)
DE Bacillus subtilis serine protease SP1 (YuxL) DNA.
XX Serine protease SP1; YuxL; detergent; ss.
XX
XX Bacillus subtilis.
XX
XX WO9903984-A2.
XX
XX 28-JAN-1999.
XX
XX 14-JUL-1998; 98WO-US14647.
XX
XX 15-JUL-1997; 97EP-0305232.
XX
XX (GEMV) GENECOR INT BV.
XX (GEMV) GENECOR INT INC.
XX Estell DA;
XX
XX WPI; 1999-132231/11.
XX P-PSDB; AAW97789.
XX
XX Use of serine protease genes from Gram-positive microorganisms - for
PT modification of host cells for the production of heterologous

PT proteins or for producing proteins for use in cleaning compositions
 XX Disclosure; Fig 1A-C; 37pp; English.

CC This DNA sequence encodes serine protease SP1 (YuxL) of *Bacillus*
 CC subtilis. Novel serine proteases SP1, SP2, SP3, SP4 and SP5 (see
 CC AA57789-93) were identified via a FASTA search of *Bacillus subtilis*
 CC genomic nucleic acid sequences. SP1 was identified by its
 CC structural homology to the S9 type serine protease dap2 of yeast.
 CC SP2, SP3, SP4 and SP5 were identified by their structural and
 CC overall amino acid homology to SP1. Host cells in which the
 CC naturally occurring gene encoding one or more of SP1, SP2, SP3,
 CC SP4 or SP5 is mutated such that the proteolytic activity is
 CC diminished or deleted altogether, can be used for the production of
 CC heterologous proteins, e.g. a hormone, enzyme, growth factor,
 CC cytokine, protease, carbohydrase, lipase, racemase, epimerase,
 CC tautomerase, mutase, transferase, kinase or phosphatase (claimed).
 CC SP1, SP2, SP3, SP4 and SP5 can also be produced on a large scale
 CC in a microbial host expression system for use in cleaning
 CC compositions such as detergents, bar or liquid soap, dish-care
 CC formulations and contact lens cleaning solutions, or for peptide
 CC hydrolysis, waste treatment, textile applications, as
 CC fusion-cleavage enzymes in protein production, and as animal feed
 CC additives.

XX Sequence 1971 BP; 558 A; 431 C; 524 G; 458 T; 0 other;

Query Match 100.0%; Score 1971; DB 20; Length 1971;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAAAAGCTGATAACCCGAGCAGATCACACGATTTGTCTGTGACCGATCCTCAA	60
DB	1	ATGAAAAGCTGATAACCCGAGCAGATCACACGATTTGTCTGTGACCGATCCTCAA	60
QY	61	TAGCCCCAGCGTACCCGCGCGATATGTAATCAACAAGTAATCAAGAAAGAT	120
DB	61	TAGCCCCAGCGTACCCGCGCGATATGTAATCAACAAGTAATCAAGAAAGAT	120
QY	121	TCGTATACATCAATATATGATCTATGAAACGAAACGGGAGGATCTGTTCTTGGACA	180
DB	121	TCGTATACATCAATATATGATCTATGAAACGAAACGGGAGGATCTGTTCTTGGACA	180
QY	181	CATGGAGAAAAGCGAACCCGACCAAGATGTTCTCCGGAGCGCGACGTTGCCCTT	240
DB	181	CATGGAGAAAAGCGAACCCGACCAAGATGTTCTCCGGAGCGCGACGTTGCCCTT	240
QY	241	ATTCTCATGAGAGCGGATCGGCACAGCTTTATATCATGACACTGAAGCGGAGAA	300
DB	241	ATTCTCATGAGAGCGGATCGGCACAGCTTTATATCATGAGACTGAAGCGGAGAA	300
QY	301	GCAAGAAACTGACTGATATCCCATATGGCGTGTCAAAGCGCTATGGTCCCGGACGGT	360
DB	301	GCAAGAAACTGACTGATATCCCATATGGCGTGTCAAAGCGCTATGGTCCCGGACGGT	360
QY	361	GAATCGATCTGCTACTATCAGTTTGGAGAGGGGAAAGCATTGATGACCGAGAAAA	420
DB	361	GAATCGATCTGCTACTATCAGTTTGGAGAGGGGAAAGCATTGATGACCGAGAAAA	420
QY	421	ACAGAGCAGACAGCTATGAACCTGTTGAAGTGCAAGGCCCTCTCTACAAACGGAGGC	480
DB	421	ACAGAGCAGACAGCTATGAACCTGTTGAAGTGCAAGGCCCTCTCTACAAACGGAGGC	480
QY	481	AAAGGGCTGACGAGAGTGCGTATGCCAGCTTGCTGTGTCAGCGTAAAGTCGGGTGAG	540
DB	481	AAAGGGCTGACGAGAGTGCGTATGCCAGCTTGCTGTGTCAGCGTAAAGTCGGGTGAG	540
QY	541	ATGAAAGAGCTGACAAAGTCAAAAGCTGATCATGGTGATCCTGTTTTCTCTGACGGC	600
DB	541	ATGAAAGAGCTGACAAAGTCAAAAGCTGATCATGGTGATCCTGTTTTCTCTGACGGC	600
QY	601	AAATGGCTGTTTTCTCAGCTAATTTAACTGAACACATGATGCCAGCAAGCCCATGAT	660
DB	601	AAATGGCTGTTTTCTCAGCTAATTTAACTGAACACATGATGCCAGCAAGCCCATGAT	660

DB	601	AAATGGCTGTTTTCTCAGCTAATTTAACTGAACACATGATGCCAGCAAGCCCATGAT	660
QY	661	GTTTACATAATGTCACCTGGAGTCTGAGATCTTTAAGCAGGTTACACCTCATCGCGGTCA	720
DB	661	GTTTACATAATGTCACCTGGAGTCTGAGATCTTTAAGCAGGTTACACCTCATCGCGGTCA	720
QY	721	TTCCGATCAAGCTCAITTTTACCAGACGGAAGTATCTTCTTTGCTTGGAAATGAAAG	780
DB	721	TTCCGATCAAGCTCAITTTTACCAGACGGAAGTATCTTCTTTGCTTGGAAATGAAAG	780
QY	781	GAATATAAGAACTGCTACGCTCTCAAGGGCTGCTATATATATCGAACAGGCGGCTC	840
DB	781	GAATATAAGAACTGCTACGCTCTCAAGGGCTGCTATATATATCGAACAGGCGGCTC	840
QY	841	ACATGCTTACTGAGATGCTGGACGTTCAATTTAGCGGATCGCTGATTTGAGATTCATTG	900
DB	841	ACATGCTTACTGAGATGCTGGACGTTCAATTTAGCGGATCGCTGATTTGAGATTCATTG	900
QY	901	ATCCGTTGGTGTGAACAGCCCGCGATTTGGACAAAGGACAGCCAAAGGTTTTATGTCATC	960
DB	901	ATCCGTTGGTGTGAACAGCCCGCGATTTGGACAAAGGACAGCCAAAGGTTTTATGTCATC	960
QY	961	GGCACAGATCAAGGCGAGTACGGCATCTATTATTTTCGATTTGAAGCCCTTGTATCCG	1020
DB	961	GGCACAGATCAAGGCGAGTACGGCATCTATTATTTTCGATTTGAAGCCCTTGTATCCG	1020
QY	1021	ATTGCTCTGAAAAAGAGTACATCAATAGCTTTTCTTTTCACTGATGAACAGCACTTT	1080
DB	1021	ATTGCTCTGAAAAAGAGTACATCAATAGCTTTTCTTTTCACTGATGAACAGCACTTT	1080
QY	1081	ATTGCCAGTGTGACAAAGCCGAGACACCGAGTGAGCTTTACAGTATCCCGCTTGGACAG	1140
DB	1081	ATTGCCAGTGTGACAAAGCCGAGACACCGAGTGAGCTTTACAGTATCCCGCTTGGACAG	1140
QY	1141	GAAGAGAAACAGTACTGCGCGGAATGACAAAGTTTGTTCAGGGAGCATACGATCAATA	1200
DB	1141	GAAGAGAAACAGTACTGCGCGGAATGACAAAGTTTGTTCAGGGAGCATACGATCAATA	1200
QY	1201	CCTGAAGAGATTCAATATGCTACAGAAAGCCGCTGATGTTGAACGCTGGCTGATGAGG	1260
DB	1201	CCTGAAGAGATTCAATATGCTACAGAAAGCCGCTGATGTTGAACGCTGGCTGATGAGG	1260
QY	1261	CCTGCACAAATGGAAGGTGAGACACATATCCACTTATTTTAACTACACGCGGCTCGG	1320
DB	1261	CCTGCACAAATGGAAGGTGAGACACATATCCACTTATTTTAACTACACGCGGCTCGG	1320
QY	1321	CATATGATCTACGGACATACATATTTTCACTGCTTTCAGTGTGCTGGCGGAAAGGATAC	1380
DB	1321	CATATGATCTACGGACATACATATTTTCACTGCTTTCAGTGTGCTGGCGGAAAGGATAC	1380
QY	1381	GGGGTCTTTTATATCAATCCGAGAGGAGCCAGGCTACGGGAGGAAATTTGTAATGCG	1440
DB	1381	GGGGTCTTTTATATCAATCCGAGAGGAGCCAGGCTACGGGAGGAAATTTGTAATGCG	1440
QY	1441	GTCAGAGGAGATTATGGGGAAGGATATGACGATGTGATGACGCTGTGGATGAGGCT	1500
DB	1441	GTCAGAGGAGATTATGGGGAAGGATATGACGATGTGATGACGCTGTGGATGAGGCT	1500
QY	1501	ATCAACAGAGATCCGATATTTGATCTTAAGCGGCTCGGTGTCACGGCGGAAAGCTACGGA	1560
DB	1501	ATCAACAGAGATCCGATATTTGATCTTAAGCGGCTCGGTGTCACGGCGGAAAGCTACGGA	1560
QY	1561	GGTTTTATACCAACTGGATCTCGGGCAGACGAAACCGCTTTTAAAGCTGCCGTTACCCAG	1620
DB	1561	GGTTTTATACCAACTGGATCTCGGGCAGACGAAACCGCTTTTAAAGCTGCCGTTACCCAG	1620
QY	1621	CGCTCGATATCAAAATGGATACGCTTTTACGGCGTCAGTGATATTCCTTTTACA	1680
DB	1621	CGCTCGATATCAAAATGGATACGCTTTTACGGCGTCAGTGATATTCCTTTTACA	1680
QY	1681	GACTGGCAGCTTTCAGCATGACATGTTTGGAGCACAGAAAGCTCTGGGACCGCTCTCT	1740
DB	1681	GACTGGCAGCTTTCAGCATGACATGTTTGGAGCACAGAAAGCTCTGGGACCGCTCTCT	1740


```
QY 901 ATCGTGGTGTGACAGCCGCCGATTTGGACAAAGGACAGCCAGGGCTTTTATGTCTATC 960
Db 886 GTCGGAGCGCTTCTCCGAAACCGCGTGGACCAAGACGGAACCGCTTTATGTCTATC 945
QY 961 GGCACAGATCAAGGACGATCGGGCATCTATTATATTCGATGAAGGCTTGTGTATCCG 1020
Db 946 GGATCTGACAGGCTTCTACAGGATTTTACTACATATCAATCAAGAGGCTTGTGTATCCG 1005
QY 1021 ATTGCTGTGGAAGAGTACATCAATAGCTTTCTCTTTTCCACCTGATGAACAGCATTT 1080
Db 1006 GTCCGGTGGAAAAGAGCATGTGACGGATTCAGCTGATCCGAGGAAAGTGGCTTT 1065
QY 1081 ATTGCCAGTGTGACAAAGCGGACAGACCGAGTGGCTTTACAGTATCC--CGTTGGA 1137
Db 1066 GCGCATCCATAGCCCTGCGCGTCCGCGGAGCAACTTTACCATCTCCAATGGTGAA 1125
QY 1138 CAGGAAGAGAAACGCTGACTGGCGGCAATGACAAGTTTGTACAGGAGCATACGATACA 1197
Db 1126 GAAAAGCGGAGCGACATAACAGATGCCAACCAAGCGTTTACGGAAGAACATCATATCG 1185
QY 1198 ATACCTGAAGAGATTCAATATGCTACAGAACAGCGGTGATGGTGAACGGCTGGCTGATG 1257
Db 1186 GAACCGGAGGAGCTTCAATTTACGCAAGCGGACGCTGACAAATTCACGGATGGCTGATA 1245
QY 1258 AGGCCTGCACAAATGGAAGTGTAGACAAACATATCCACTTATTTTAAACATACACGGCGT 1317
Db 1246 AAGCCGCCCAATATGAAAAGAAACACAGTATCCGCTCATTTCTGGAAGTGCACGGCGGA 1305
QY 1318 CCGATATGATGTACGACATACATATTTTCATAGTTTCAAGTGTCTGGCGGCAAGGA 1377
Db 1306 CCTCACCCCATATGATGCAATGCTTATTTTCATGAATTTCAAGTGTCTGGCGCAAGGGA 1365
QY 1378 TAGCGGTGCTTTATCAATCCAGAGGAAGCACAGGCTACGGGACGGAATTTGTGAAT 1437
Db 1366 AGCCCGCTGTATGTAAACCCCGGGAAGCCACGGCTACGACAGATTTGTCAAC 1425
QY 1438 GCGGTACAGAGGATTTATGGGGAAGGATTTATGACGATGTGATGACGGCTGTGGATGAG 1497
Db 1426 CGGCTGAGGGCGCATACGGCGGTGGAGATTTTAAGGATGTAAATGGCGCTTGTACCCAC 1485
QY 1498 GCTATCAACAGATCCGCATATTTGATCCTAAGCGGCTCGTGTACAGGGCGGAAGCTAC 1557
Db 1486 GTGCTTGAACATTAATGATTTGCTGCATCAAGAAAGGCTCGGCATCAACGGAGAGCTAC 1545
QY 1558 GGAGGTTTATGACCAACTGGATCGTGGGACAGACGACCGCTTTAAAGCTGCCGTACC 1617
Db 1546 GGGGCTTTATGACAACTGGCTGTGGCCACAGAGCGCTTCAAGCCCGCATGTACA 1605
QY 1618 CAGGCTCGATATCAAAATGGATCAGCTTTTACGGCGTCACTGATATCGGCTATTTCTTT 1677
Db 1606 CAGAGGTCCATTTTCCAACTGGATCAGTGTTTTACGGGTAAAGCGATCGGCTATTTTTC 1665
QY 1678 ACAGACTGGCAGCTTGACATACATGTTTTCAGGACACAGAAAGCTCTGGGACCGGCT 1737
Db 1666 ACAGACTGGCAGCTTGACATACATGTTTTCAGGACACAGAAAGCTCTGGGACCGGCT 1725
QY 1738 CCTTAAATACGCAAGCTGGAGACACCGCTTTTGTACTGATGCGAGCGGAGAT 1797
Db 1726 CCGCTCAAAATACGCGGACAGGTGGAGACCCCTTCTCATCTGCTATGCGGACGGAC 1785
QY 1798 GACCGATGCCGATCGAGCGGAGGACGCTGTTTATCGCTCTGAAAAAATGGGCAAG 1857
Db 1786 GACAGGTGCCGATTAACAGCGGGAAGCTGTTTACGGCGCTGAAAAAATGGGAAA 1845
QY 1858 GAACCAAGCTGTGCTGCTTTCGAATGCATCGCACATTTTATCACGACCGGACACCA 1917
Db 1846 GAAGTAAATCTGTCAGATTTTCCGAACGCTGCACAGATTTTATCAAGAAGCGGACCG 1905
QY 1918 AGACAGCGGATCAAGCGCTGAATTTATATACAGCTCATGTTTGTATCAACATCT 1970
Db 1906 AAGCAGCGGATCAGCGGCTTGACTATATTTCAGGCTGGTTTGAAGCTTATCT 1958
```

```
RESULT 3
AAH41227/c
ID AAH41227 standard; DNA; 265118 BP.
XX AAH41227;
XX 29-OCT-2001 (first entry)
DE Pyrococcus abyssi genomic fragment #6.
KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX Pyrococcus abyssi.
XX OS Pyrococcus abyssi.
XX FH Key Location/Qualifiers
FT misc_feature 1..49980
FT /tag= a
FT /note= "This sequence overlaps with the 3' end of
FT AAH41226"
XX PN FR2792651-Al.
XX PD 27-OCT-2000.
XX PF 21-APR-1999; 99FR-0005034.
XX PR 21-APR-1999; 99FR-0005034.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX PI Porterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Quetrou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX Claim 1; Page 593-665; 1657pp; French.
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P.
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAB41226.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade.
XX Note: This patent is in the same patent family as WO200065062, which
XX contains additional sequences as shown in AAB99132-AAB99143,
XX AAH75903-AAH75920 and AAG66436.
XX
XX Sequence 265118 BP; 75226 A; 61302 C; 54005 G; 74585 T; 0 other;
XX
XX Query Match 12.1%; Score 238.2; DB 22; Length 265118;
XX Best Local Similarity 55.3%; Pred. No. 7.5e-64;
XX Matches 482; Conservative 0; Mismatches 386; Indels 3; Gaps 1;
```

```
QY 1091 TGACAAAGCGGACACACCGAGTTCAGCTTTACAGTATCCGCTGGACGGAAGAGAAC 1150
Db 58851 TTACAGCTCAAGATGGCGACAACTCCGACTGAGCTTTACGCTTCAGGGATGGAAAGAGA 58792
QY 1151 AGCTGACTGGCGGAATGACAAGTTTGTACGGGAGCATACGATATCAATACCTGAAGAGA 1210
Db 58791 AGCGTATCAGACTTTACGCTGGATTTAGGACTATAACTTTCAAACCCAGGCACT 58732
QY 1211 TTCAATATGCTACAGAGAGCGGCTGATGTGAACGGCTGGCTGATGAGGCTGCACAAA 1270
Db 58731 TCAGGGTTACGGCTAGCGATGGGAAGAAATAGATGCTGATCATGAAGCCCGTTGACT 58672
QY 1271 TGAAGGTGAGACAAACATATCCACTTATTTTAAACATACAGCGGCGCTCGCATATGATGT 1330
```

```
Db 58671 TTAGCTGGAGGAAGTATCTCTGCACTGCTGGAATTCACGGTGGCCAAAACACACCT 58612
Qy 1331 ACGGACATACATATTTTCATGAGTTTCAGTGTCTGGCGGCGAAAGGATACGGGTCGTTT 1390
Db 58611 ACGGTTATGATTTATGACGAGTTTCAGGTTTATGTTGCGAAAGGCTTCGTTGTAATCT 58552
Qy 1391 ATATCAATCCGAGGGAAGCCACGGCTACGGCAGGAAATTTGTCAATGCGGTGAGAG 1450
Db 58551 TTTCAATCCAGGGGAAGTGATGGTTATGGTGAGGAGTTTCGGC---GACATAGGGGTC 58495
Qy 1451 ATTATGGGGAAGGATTTATGACGATGTGATGACGGCTGTGGATGAGGCTATCAACGAG 1510
Db 58494 ACTACGGTGAAGCGGATTTATCAAGACCTAATGAGGTAGTTGACGAGGCTTTAAAGCCT 58435
Qy 1511 ATCCGATATTTATGCTTAAGCGGCTCGGTGTACGGCGGGAAGCTACGGAGGTTTATGA 1570
Db 58434 TTGACTTTATAGATGAGAAAGGCTTGGCGTTACCGGAGGATCATACGGAGGCTTCATGA 58375
Qy 1571 CCAACTGGATCGTCGGGCAGACGACGCTTTTAAAGCTGCGGTTACCCAGCGCTCGATAT 1630
Db 58374 CAAATGGATCGTTGGGCACACAAACAGATTCAAGCTGCGGTAACCTACGCGCTCAATAT 58315
Qy 1631 CAAATGGATCAGCTTTACGGGCTCAGTGATATCGGCTATTTCTTTACAGACTGGCAGC 1690
Db 58314 CCAACTGGATCAGCTTCTTTTGAAGACGACGATATCGGCTATTTACTCGCTCCAGATCAGA 58255
Qy 1691 TTGAGCATGATGTTTGGAGACACAGAAAGCTCTGGGACCGGCTCTCTTTTAAATACG 1750
Db 58254 TCGGTGGGATCTCTTGGAGTAACCTAGAGGCTACTGGGAGAGAGGCCATTTGAAGTACG 58195
Qy 1751 CAGCAACGTTGGAGACACCGCTTTTGATCTGATGCGGAGGATGACCGGATGCCCGA 1810
Db 58194 CTCAACGTTGAGACTCCGCTCTGATATTTACTCCAGGAGGACTACGCTGTGGC 58135
Qy 1811 TCGAGCAGCGGAGCAGCTGTTATCGCTCTGAAAAAATTTGGGCAAGGAACCAAGCTTG 1870
Db 58134 TTCCGAGGCTTTGCAATTTTACATAGCGTTGAAGTATTTTGGGCAACCGTCGAGCTGG 58075
Qy 1871 TCGGTTTCCGAATGCATCCACAAATTTATCAGCACCAGGACCCCAAGACGGGATCA 1930
Db 58074 CGATATTCAGGTTGAGATGATGATGAGTGGAGGTGGAAGCCCAAGCATAGGGTTA 58015
Qy 1931 AGCGCTGGAATTTATCAGCTCATGTTTGA 1961
Db 58014 AGAGCTCAGCTCATAGTGGGTGTTGA 57984
```

```
RESULT 4
ID AAV33582
XX AAV33582 standard; DNA; 1896 BP.
AC AAV33582;
XX
XX
XX
XX 29-DEC-1998 (first entry)
XX Pyrococcus horikoshi acylpeptide hydrolase encoding DNA sequence.
XX
XX Pyrococcus horikoshi acylpeptide hydrolase; ss.
XX
XX Pyrococcus horikoshi.
OS
XX Location/Qualifiers
FH 1..1896
FT /*tag= a
FT /product= "Pyrococcus horikoshi acylpeptide hydrolase"
FT /note= "CDS does not contain a stop codon"
XX
XX JPI0210977-A.
XX
XX 11-AUG-1998.
XX
XX 31-JAN-1997; 97JP-0018381.
```

```
XX 31-JAN-1997; 97JP-0018381.
XX (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX WPI: 1998-488369/42.
XX P-PSDB; AAW70508.
XX A heat-resistant acylpeptide hydrolase and a gene coding it - useful
XX for hydrolysing the C termini of proteins at high temperature
XX Example 5; Pages 6-7; 8pp; Japanese.
XX The present sequence represents a Pyrococcus horikoshi acylpeptide
XX hydrolase encoding DNA sequence. The enzyme has the following
XX properties: (a) it hydrolyses an acylpeptide; (b) an optimum temperature
XX of 90-95 degrees Centigrade; (c) an optimum pH of 5.0-6.0; (d) no loss
XX of activity when heated to 95 degrees Centigrade for 3 hours at pH 7.5;
XX and (e) a molecular weight of 60 kDa. The enzyme is claimed useful
XX for hydrolysing the amino end of an acylated protein and a peptide at
XX high temperatures.
XX Sequence 1896 BP; 614 A; 358 C; 491 G; 433 T; 0 other;
XX
XX Query Match 11.5%; Score 227.2; DB 19; Length 1896;
XX Best Local Similarity 55.4%; Pred. No. 1.2e-61;
XX Matches 461; Conservative 0; Mismatches 368; Indels 3; Gaps 1;
Qy 1139 AGGAAGAAACACAGCTGACTGGCGGAATGACAAAGTTTGTTCAGGAGCATACGATATCAA 1198
Db 1061 ATGCAAGGAGGAAGGTTACCACCTTTTAAACAATGATAAAGGTTACACCCCTTTCAA 1120
Qy 1199 TACCTGAAGAGATTCAATATGCTACAGAGAGCGGCTGATGGTGAACGGCTGGCTGATGA 1258
Db 1121 AACCTGAACACTTTAAAGGTTAAAGCAAGTACGGGGTTGAAATAGATGCTGGGTAATGA 1180
Qy 1259 GGCCTGCACAAATGGAAGGTGAGACACATATCCACTTTATCTTAAACATACACGGCGTC 1318
Db 1181 AACCGTGAACCTCAGGAAAGGAAGATATCCAGCTATTCTAGAGATCCACGGTGGTC 1240
Qy 1319 CGCATATGATGACGGACATACATATTTTCATAGATTTTCAGTGTCTGGCGGGAAGAT 1378
Db 1241 CTAAACCGCTTACGGTTACGCTTTTATGACAGGTTTCCAGGTTTAAAGCTTAAAGGCT 1300
Qy 1379 ACGCGTCTGTTTATATCAATCCGAGAGAACCCAGGCTACGGCAGCAATTTGTAATG 1438
Db 1301 TCGCTGATATTTCAAACTCTAGAGGAGCGATGGCTACGGAGAGGTTTCGCGGAT- 1359
Qy 1439 CGGTGAGAGGATTTATGGGGAAGGATTTATGACGATGTGATGACGCTGTGGATGAGG 1498
Db 1360 --ATAAGGGGACACTATGGGAGAGGATTTACCAGGATTTAATGAGGTAGTCGATGAAG 1417
Qy 1499 CTATCAACAGAGATCCGCATATTTGATCTAAGCGGCTCGGTGTCTACGGCGGAAAGTACG 1558
Db 1418 CATTAAGGAGATTTGACTTTCATAGATGGGAAAGGCTAGGAGTTACCGGGGTTCTCTATG 1477
Qy 1559 GAGGTTTATGACCAACTGGATCTCGGCGACAGCAACCGCTTTTAAAGCTCCGCTTACCC 1618
Db 1478 GTGGCTTCATGACGAACCTGGATAGTCGGACATACCAACAGGTTCAAGCCGCTGTAACCC 1537
Qy 1619 AGCGCTCGATATCAAAATTTGATCAGCTTTTCACGGCGTTCAGTGATATCGGCTATTTCTTTA 1678
Db 1538 AGAGATCAATTTCAAAATTTGATAGCTTCTTCGGGACAAACGATATAGTTTATCTTTG 1597
Qy 1679 CAGACTGGCAGCTTGAGCATGACATGTTTGGAGGACACAGAAAGCTCTGGGACCGGCTCTC 1738
Db 1598 CTCAGATCAAAATAGGAAAGATCCCTGGAGCACTTGGAAAGGTTATTGGGAAAAAGAGCC 1657
Qy 1739 CTTTAAATACGACAGCAACCTGGAGACACCGCTTTTTCATCTACTGCATGCGGAGCGGATG 1798
Db 1658 CATTAAGTACGCTCCCAACGTTGAAACTCCCGCTGCTTATATCCACTCTACCGAAGAT 1717
Qy 1799 ACCGATGCCGATCGAGCAGCGGAGGAGCTGTTTATCGCTCTGAAAAAATTTGGCAAGG 1858
```

Db 1718 ACAGGTGTTGGCTCCCGAGGCAATTGCAACTCTTCATATCCCTAAATACCTGGGAAGA 1777
 QY 1859 AAACCAAGCTGTCCGTTTCCGAAATCATCGCAATTTATCATCGACCGACCCCAA 1918
 Db 1778 GAGTTGAATGGCAATATCCAGGAGAAATCATGACCTAAGTAGATCTCTGGGAAGCCAA 1837
 QY 1919 GACAGGGGATCAAGCGCTGAATATATATCATGCTCATGTTGTGATCAACATCT 1970
 Db 1838 AGCACAGGGTTAAAGACTTGAACATAATAGCAGGATGGATGAGAAATGGCT 1889

RESULT 5

ABQ86198

ID ABQ86198 standard; DNA; 1032 BP.

AC ABQ86198;

XX 09-SEP-2002 (first entry)

XX Lactobacillus rhamnosus HN001 polynucleotide SEQ ID NO 14.

KW Lactobacillus rhamnosus; strain HN001; vulnery; antilipaeamic; milk;
 immunostimulant; anti-infection; lactose digestion; immune system;
 metabolic activity; nutrition; health; transgenic; lactic acid bacteria;
 genome mapping; gastrointestinal disorder; dairy processing; vaccine;
 fermentation; probiotic; cholesterol; wound healing; gene; ds.

XX Lactobacillus rhamnosus.

XX W0200244383-A1.

XX 06-JUN-2002.

XX 28-NOV-2001; 2001WO-NZ00286.

XX 28-NOV-2000; 2000US-0724623.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.

XX Glenn M, Havukkala IJ, Lubbers MW, Dekker J;

XX WPI; 2002-519588/55.

XX P-PSDB; ABP61033.

PT Novel isolated Lactobacillus rhamnosus polynucleotides encoding
 PT polypeptide with anti-infection/lactose digestion modulating activity,
 PT useful to improve properties of microbes used in milk-derived products
 PT manufacture

XX Claim 2; Page 63; 128pp; English.

XX The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence (ABQ86185-ABQ86243) present in Lactobacillus
 CC rhamnosus strain HN001 that encodes a polypeptide (ABP61020-ABP61060)
 CC with activity such as enzyme activity; anti-infection activity; lactose
 CC digestion modulating activity; immune system modulating activity; amino
 CC acid, lipid, vitamin or carbohydrate metabolic activity, flavour, texture
 CC or aroma modulating activity. (I) is useful for improving the properties
 CC of microbes used in the manufacture of milk-derived products and
 CC probiotic supplements, which involves modulating the polynucleotide
 CC content or composition of the microbes by transforming the microbes with
 CC (I). (I) is also useful for identifying an organism (preferably a
 CC bacterial or yeast cell) or reproductive material or an extract from the
 CC organism, as having a specific origin. Proteins encoded by (I) are useful
 CC for modifying the flavour, aroma, texture and/or nutritional and health
 CC benefits of milk-derived products, which involves adding one or more
 CC polypeptides to the milk being processed. Genetic constructs comprising
 CC (I) are useful for modulating the polynucleotide content or composition
 CC of a organism. (I) is useful for identifying, isolating or synthesising
 CC DNA molecules such as promoter, DNA binding elements, open reading frames
 CC or full-length genes, that then can be used as expressible DNA in

CC transgenic organisms. (I) may be used to detect lactic acid bacteria,
 CC preferably L. rhamnosus in a sample material. (I) is also useful for
 CC genome mapping, physical mapping, and in positional cloning of genes of
 CC more or less related microbes, and to design probes and primers. (I) is
 CC also useful for transforming microbes for use in a therapeutic
 CC composition that is effective for treating or preventing a
 CC gastrointestinal condition or disorder caused by the presence of
 CC pathogenic microbes in the gastrointestinal tract or by the absence of
 CC normal intestinal microbes in the intestinal tract. Proteins are used to
 CC raise antibodies, to isolate corresponding interacting proteins, as
 CC nutritional additives and as additives in dairy processing and
 CC fermentation processing. (I) and encoded proteins are used for the
 CC selection and production of more effective probiotic bacteria, as
 CC bioactive (health promoting) ingredients and health supplements, for
 CC immune function enhancement; for reduction of blood lipids such as
 CC cholesterol; for production of bioactive material from genetically
 CC modified bacteria as adjuvants; for wound healing; in vaccine
 CC development, in selection and production of genetically modified rumen
 CC microorganisms for improved animal nutrition and productivity, better
 CC flavour and improved milk composition.

XX Sequence 1032 BP; 252 A; 220 C; 281 G; 279 T; 0 other;

Query Match

9.9%; Score 195.6; DB 24; Length 1032;

Best Local Similarity 54.3%; Pred. No. 1.1e-51;

Matches 469; Conservative 0; Mismatches 379; Indels 16; Gaps 3;

QY 1050 CTTTCTCTTTCACCTGATGACAGCACCTTTATTTCCAGTGTGACAAAGCCGGACAGACC 1109

Db 78 CTTTGCATTGACACCAGATCATCGTGGGTGGTTTCTACTGAGAGTACGATGACCATTC 137

QY 1110 GAGTGAGCTTTACAGTATCCCGCTTGGACAGGAGAGAAACAGCTACTGGCGGAATGA 1169

Db 138 GAGTCGGCTGGTTTATTTTGGATTTGGCATCGGAAGAGACAGGTTTGTATGACCCGAA 197

QY 1170 CAAGTTTGTGAGGAGCATACGATATCAATACCTGAAGAGATTCAATATGCTACAGAAGA 1229

Db 198 TCGTCAGGTAACACAGCTCACTTGGGCTTAGTTACCCCTCAACAGCTTAAATTTTCAACGAGA 257

QY 1230 CGGCGTGATGTTGAACGGCTGGCTGATGAGGCTGCACAAATGGAAGTGCAGACAAATA 1289

Db 258 CGGTTTGTAGATTGAGGCTGGTATTTTCCACCGCAACAGCGGTCA-----TCATCGCA 311

QY 1290 TCCACTTATTTTAAATACACAGCGGCTCCGATATGATGATGAGGATACATATATTTTCA 1349

Db 312 TCCGCAATTTTGTATGTCATGTCGCGCCAGCAGTCGGATATGGCTATACCTTTTCCA 371

QY 1350 TGAGTTTCAAGTGTGGCGGGAAGGATACGCGGTGCTTTATATCAATCGGAGGAGAG 1409

Db 372 TGAATGCAATCTCTGGCAGCAAAAGGCTATGCGGTGATTTGTGCAAAATCCGCTGGAGG 431

QY 1410 CCACGGCTACGGCGAGGAATTTGTGAATGCGGTGAGAGGATTTATGGGGAAGGATTA 1469

Db 432 GTTAGTTACCGGAGGCAATTTACGGCGCTGTCATTAACATTAACCGCGCGGATTA 491

QY 1470 TGACGATGTGATGCGGCTGTGGATGAGGCTATCAACAGAGATCCGATATGATCCTAA 1529

Db 492 TGAAGATTGCTTGGCTTCGGGTGAAGAAGCGCTAAAGCTCGATCAACAATATGATCCGCA 551

QY 1530 GGGGCTCGGTGTCACGGCGGAGCTACGGAGGTTTTATGACCAACTGGATTCGCGGCA 1589

Db 552 AGCTTATTTGCTACCTGGCGGTTCTTATGGCGGGTTTATGACTGAATGTTGACCCCA 611

QY 1590 GACGAACCGCTTTAAAGCTGCGGTTTACCCAGCGCTCGATATCAAAATGATCATCTTTCA 1649

Db 612 TACGATCGTTTTAAAGCAGCGGTACCCAGCGTTCGATTTCCAAATGGCTGAGTATGA 671

QY 1650 CGGCGCTAGTATATCGGCTATTCTTTACAGACTGGCAGCTTGA-----GCATGA 1700

Db 672 TGGTACCAGTACATCGGTTATTTACTTTACCGCTGGGAAGTGAAGAAAGTGGACTGG 731

QY 1701 CATGTTTGGAGCACAGAAAAGCTCTGGGACCGGCTCTCTTTAAATACCCAGCAACGT 1760

Db 732 CGATTTGTCAGATGTGCAAGCGCTTTGGGATTTTTCACCATTTAGTCACATTTGATCATGC 791
QY 1761 GGAGACACCGCTTTTGTACTGATGCGGAGCGGGATGACCGATGCCCGATC-GAGCAGG 1819
Db 792 CAGAACACCGACGCTTGTGTGATGACACATGAAATGATGAACGCTGCCCGCATCGGCCAAG 851
QY 1820 CGGAGCAGCTGTTTATCGCTCTGAAATAAATGGCAAGGAAACCAAGCTTGTCCGTTTTC 1879
Db 852 TAGAAGTTGATCATCGGCTCAAACTGCATGCTGTGTAACCAAGTTTCATCGGTTTC 911
QY 1880 CGAATGCATCGCACAATTTATCAC 1903
Db 912 CAAAGTCAATCATGATTTGTCCC 935

RESULT 6

ABK77575
ID ABK77575 standard; DNA; 477 BP.

XX ABK77575;

XX 13-AUG-2002 (first entry)

XX Bacillus clausii genomic sequence tag (GST) #418.

XX Differential gene expression; genomic sequenced tag; GST;

XX altered culture condition; environmental stress;

XX physiological provocation; ds.

XX Bacillus clausii.

XX W0200229113-A2.

XX 11-APR-2002.

XX 05-OCT-2001; 2001WO-US31437.

XX 06-OCT-2000; 2000US-0680598.

XX 27-MAR-2001; 2001US-279526P.

XX (NOVO) NOVOZYMES BIOTECH INC.

XX (NOVO) NOVOZYMES AS.

XX Berka R, Clausen IG;

XX WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -

XX Claim 11; SEQ ID NO 4866; 200pp; English.

XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 477 BP; 138 A; 89 C; 118 G; 132 T; 0 other;

Query Match 9.8%; Score 192.2; DB 24; Length 477;
Best Local Similarity 65.1%; Pred. No. 8.6e-51;
Matches 300; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

QY 1499 CTATCAAAACGAGATCCGCATATTGATCCTTAAGCGCTCGGTGTACACGGCGCAAGCTACG 1558
Db 7 CAATTGCAACAATAGCTGGATTGATCAGGAGCGCATCGGTCTGCTGGGGGAAGTTACG 66

QY 1559 GAGTTTTATGACCAACTGGATCTCGGGGAGAGAACCGCTTTAAAGCTCGGTTACCC 1618
Db 67 GTGGCTTTATGACGAATTGGGCTGTGGCCATAGCAATCGGTTCAAAGCGCGCTTACGC 126

QY 1619 AGCGCTGATATCAAAATGGATCAGCTTTCACGGCGTCAGTCATATCGGCTATTCTTTA 1678
Db 127 AAGCTCGATATCAAAATGGATTAGTTTTATGGCGTAAGCGATATTGGCTACTATTTTT 186

QY 1679 CAGACTGGCAGCTTGAGCATGACATGTTTGAGGACACAGAAAAGCTCTGGACCGGTCTC 1738

Db 187 CAGAGTGGCAAAATAAAGACAGAC---TTGCATGATATCGAAACACTGTGGCCCATTCGC 243

QY 1739 CTTTAAATACGACAGCAAAAGCTGGAGACACCGCTTTTGATACTGCATGGCGAGCGGATG 1798

Db 244 CACTCAAAATACGTTGAAAAGCTGGAACGCGCATTTGCTTATTCTGCACGCTGAAAAGATT 303

QY 1799 ACCGATCCCGCATCGACAGCGGAGCAGCTGTTTATCGCTCTGAAAATAATGGCAAG 1858

Db 304 ATCGCTCTCGGATTGAAACAGCAGACGAAATGTTTATTCGTTGAAAAGCATGGGAAG 363

QY 1859 AAACCAAGCTTGTCGGTTTTCGGAATGCATCGCACAAATTTATCAGCACCGACACCCAA 1918

Db 364 AGACTGTATTATTCCTCGTTTCCAGAGCAATCATGAACATCTCAAGAAGTGTAAACCTA 423

QY 1919 GACAGCGGATCAAGCGCTGAAATATATATACATCATGCTT 1959

Db 424 ATTTGCGGATTGACGCTCTCAATCGCATAGCAGATTGTTT 464

RESULT 7

AAH41226/C

ID AAH41226 standard; DNA; 349980 BP.

XX AC AAH41226;

XX DT 29-OCT-2001 (first entry)

XX Pyrococcus abyssi genomic fragment #5.

XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.

XX Pyrococcus abyssi.

XX Key Location/Qualifiers

FT misc_feature 1..49980

FT /*tag= a

FT /note= "This sequence overlaps with the 3' end of

FT AAH41225"

FT 300001..349980

FT /*tag= b

FT /note= "This sequence overlaps with the 5' end of

FT AAH41227"

XX FR2792651-A1.

XX 27-OCT-2000.

XX 21-APR-1999;

XX 99FR-0005034.

[illegible]

Db	6328	YMTRRSVGYRARGVARTTYYRRRGYGYRGYGVGNHNHTYRGVNGYRRAARSRGNAAG	6387
QY	1515	GCATATTGATCTCTAAAGCGGCTCGGTGTCACGGCGGGAAGCTACGGAGGCTTTTATGACCAA	1574
Db	6388	YHGYYTRGNGNGNVARGYMTARGRGYAAAAARVARASHHHVARTVAGNGYGVNGNGYGN	6447
QY	1575	CTGGATCGTCGGCAGACGAAACCGCTTTAAAGCTGCGGTTACCGAGCGGCTGCATATCAAA	1634
Db	6448	ARGRGYARGARGAAVAGNCGNRGNRVARTMTRGNGNMTFRAGGYARGVATYRARG	6507
QY	1635	TTGGATCAGCTTTCACGGCGCTCAGTCGATCGCGCTATTCTCTTACAGACTGGCAGCTTGA	1694
Db	6508	TYRRRGYRGYMTRASVARMTTGGYVAAAGGYGTHSRVARTRASVAGYGYMTRVRAARG	6567
QY	1695	GCATGATCATTTTGAGGACACAGAAAGCTCTCGGACCGGCTCTCCTTTAAAAATACGCAGC	1754
Db	6568	ASAASTRGNGNRGYAAAAAPRAAAASNAASRRGGNGNARGTHMTGYGASNYTYYRVVAGG	6627
QY	1755	AAACGTGGGACACCGCTTTTGTGATCTGCATGGCGAGCGGGATGACCGATGC	1806
Db	6628	NGRASASNAAAAYSVATHRGYMTGTMATAGNTHRGVVAHSGSGRRGAAYSAAYSV	6679

RESULT 13	
AAH24065/c	
ID	AAH24065 standard; DNA; 4590 BP.
XX	
XX	AAH24065;
AC	
XX	
XX	
DT	29-AUG-2001 (first entry)
XX	
XX	
DE	Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
XX	
XX	Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
KW	modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
KW	lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
KW	functional food; transgenic yeast; fat/lean ratio; food use; ds.
XX	
XX	Saccharomycetes cerevisiae.
OS	

xx	Key	Location/Qualifiers
FT	misc_feature	10
FT		/*tag= a
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3617
FT		/*tag= b
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3649
FT		/*tag= c
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3679
FT		/*tag= d
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3819
FT		/*tag= e
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3862
FT		/*tag= f
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3864
FT		/*tag= g
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3888
FT		/*tag= h
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3890
FT		/*tag= i
FT		/*note= "Represented as * in the specification"
FT	misc_feature	3912
FT		/*tag= j
FT		/*note= "Represented as * in the specification"
FT	misc feature	3914

FT		/**tag= k	"Represented as * in the specification"
FT		/note= 3938	
FT	misc_feature	/**tag= l	"Represented as * in the specification"
FT		/note= 3939	
FT	misc_feature	/**tag= m	"Represented as * in the specification"
FT		/note= 3941	
FT	misc_feature	/**tag= o	"Represented as * in the specification"
FT		/note= 3943	
FT	misc_feature	/**tag= p	"Represented as * in the specification"
FT		/note= 4361	
FT	misc_feature	/**tag= q	"Represented as * in the specification"
FT		/note=	
XX	WO200133977-A1.		
PN	17-MAY-2001.		
PD			
XX			
XX	06-NOV-2000; 2000WO-AU01362.		
PP			
XX			
PR	05-NOV-1999; 99AU-0003875.		
XX			
PA	(META-) METABOLIC PHARM LTD.		
XX			
PI	Belyea CI, Ng FM, Vaughan P;		
XX			
DR	WPI; 2001-328876/34.		
XX			
XX	New organisms containing nucleic acid encoding a growth hormone		
PT	fragment which modulates lipid metabolism are useful to produce dietary		
PT	aids for obesity and in the meat production industry -		
XX			
PS	Disclosure; Page 48-50; 54pp; English.		
XX			
CC	The invention relates to novel transgenic organisms useful in the		
CC	production of functional food and drink products for the treatment		
CC	or prevention of obesity via the regulation of lipid metabolism. The		
CC	organisms comprise a polynucleotide encoding a growth hormone fragment		
CC	capable of stimulating the activity of hormone-sensitive lipase (the key		
CC	enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key		
CC	enzyme in lipogenesis). The growth hormone fragment preferably contains		
CC	at least the disulphide-bonded loop of a mammalian growth hormone (but is		
CC	not the full-length growth hormone) and is optionally linked to an		
CC	epitope tag or heterologous fusion protein partner. The transgenic		
CC	organism may be a microorganism used to produce a fermented product		
CC	(e.g., yeast), or an edible plant or animal or cell thereof. Food or		
CC	drink made using methods of the invention are used to modify fat/lean		
CC	ratio, lipid metabolism or food use in a mammal. In particular, the food		
CC	or drink products may be used to treat or prevent obesity, particularly		
CC	in humans, and may also be used to improve the fat/lean ration of		
CC	livestock raised for meat production. In the exemplification of the		
CC	invention, the human growth hormone (hGH) fragment analogue AOD9604 was		
CC	expressed in yeast, optionally fused to the FLAG epitope (AAB73625).		
CC	The present sequence is described as a DNA sequence from yeast in		
CC	the sequence listing, but is not further referred to in the		
CC	specification.		
XX			
SQ	Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;		
	Query Match 2.0%; Score 38.6; DB 22; Length 4590;		
	Best Local Similarity 10.9%; Pred. No. 0.72;		
	Matches 79; Conservative 323; Mismatches 322; Indels 4; Gaps 1		
QY	661 GTTTCATATAGTCACCTGGAGCTTGAGCATCTTAAGCAGGTACACCCTCATCGGCTCA 720		
	:: :: : : : : : : : : : : : : : : : : : : : :		
Db	4540 GYHKASRSABYSGYTHCNNTYACNNTYAGVVTDTSATGASKSYSSVKGPHGARXGSHYA 448		
	:: :: : : : : : : : : : : : : : : : : : : : :		
QY	721 TTCGGATCAGGCTCATTTTCACACGACGGAAGGTATCTTGCTTGGTAATGAAAG 780		
	:: :: : : : : : : : : : : : : : : : : : : : :		

	Query Match	2.0%;	Score 38.6;	DB 22;	Length 4590;
	Best Local Similarity	10.9%;	pred. No. 0.72;		
	Matches	79;	Conservative 323;	Mismatches 322;	Indels 4; Gaps 1;
y	661	GTTCATATATGTCACCTGGAGATCTTAAAGCAGGTACACCTCATCGCGCTCA	720		
b	4540	GYHKRSABSYGTHCNNTYACNNTYAGVYTDATGASKSYSVKGTGKARYGSHYNA	4481		
y	721	TTCCGGATCAAGCTCATTTTCACCAAGCGAAGGTATCTTGCTTTGCTTGGAAATGAAAG	780		

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX SQ Sequence 788 BP; 91 A; 86 C; 291 G; 320 T; 0 other;

Query Match	1.9%;	Score 36.6;	DB 24;	Length 788;
Best Local Similarity	49.2%;	Pred. NO. 1.1;		
Matches	96;	Conservative	0;	Mismatches 99; Indels 0; Gaps 0;
Qy	49	ACCGATCCTCAATACGCCCGCCAGACGGTACCCGTCGCCGATATGTAAATCACAAGTAAT	108	
Db	264	AACGAAACGGCCCGCCCACTAACAACTCGACCGCTACTCAATCAACAAAAA	205	
Qy	109	CAAGAGAAAGATTGCTATACATCAATATATGATCTATGAACGAAACGGGAGGATCT	168	
Db	204	AAACTCGNAACCGCTACGAAACAAATACCGAACTACGAAACAAACCGAACCGAA	145	
Qy	169	GTTCCTTGGACATGGAGAAAGCGAAGCAGCCAGCCCAAGATGCTCTCCGACGGGCGC	228	
Db	144	ACGCCGTAAAAAAGAAAAAAGCGAAATCTTCCCAAAAAAAGTCCGCCGCGCCCGA	85	
Qy	229	ACGCTTGCCTTATT	243	
Db	84	ACGACCGCCTTAAT	70	

Search completed: July 4, 2003, 13:20:01
Job time : 458 secs

Db 1301 TCCTCGTGATATCTCAATCTTAGAGGCGGATGGCTACGGAGGAGTTCCGGGAT- 1359
QY 1439 CGGTGAGGAGATATGGGGAAAGATATGACGATGTGATGACGCTGTGATGAGG 1498
Db 1360 --ATAAGGGACATATGGGAGAGGATACCAAGGATTTAATGAGGTAGTCGATGAAG 1417
QY 1499 CTATCAACAGATCCCGCATATGATCTTAAGCGGCTCGGTGTCACGGCGGAAGCTAGG 1558
Db 1418 CATTAAGGAGATTTGACTTCATAGATGGGAAAGCTAGGAGTTACCGGGGGTTCCTATG 1477
QY 1559 GAGGTTTTATGACCAACTGGATCGTCGGGACAGAAACCGCTTTAAAGCTGCCGTTACCC 1618
Db 1478 GTGGCTTCATGACGAAGTGGATAGTCGACATACCAACAGGTTCAAAAGCCGCTGTAAACC 1537
QY 1619 AGCGCTCGATATCAAAATGATCAGCTTTACGGGTCAGTGATATCGGCTATTTCTTTA 1678
Db 1538 AGAGATCAATTTCAAAATGGTAAGCTTCTTCGGGACAAACGATATAGTTATTACTTTG 1597
QY 1679 CAGACTGGAGCTTGAGCATGACATGTTTGGAGGACACAGAAAGCTCTGGGACCGGTCTC 1738
Db 1598 CTCAGATCAATAGGAAAGATCCCTGGAGCACTTGGAGGTTATTGGGAAAGAGCC 1657
QY 1739 CTTTAAATAGCAGCAACGTGGAGACACCGCTTTTATGATGATGCGGAGCGGGATG 1798
Db 1658 CATTAAGTACGCTCCCAACGTTGAAACTCCCTGCTTATAATCCACTCTACCGAAGACT 1717
QY 1799 ACCGATCCGATCGAGCAGCGGAGCGAGCTGTTATCGCTCTGAAAAAATGGCAAGG 1858
Db 1718 ACAGGTGTGCTTCCCGAGGCAATGCAACTCTTCATATCCCTAAATACCTGGGGAAGA 1777
QY 1859 AAACCAAGCTTGTCGGTTTTCCGAATGCATCGCACAAATTTATACGCAACCGGACACCCAA 1918
Db 1778 GAGTTGAATGGCAATATCCAGGAGAAATCATGACCTAAGTAGATCTGGGAAGCCAA 1837
QY 1919 GACAGCGGATCAAGCGCTGAATATATACGCTCATGTTGATCAACATCT 1970
Db 1838 AGCAGAGGTTAAAGACTTGAACCTAATAGCAGGATGATGGAGAAATGGCT 1889

RESULT 2

US-08-664-646A-1

; Sequence 1, Application US/08664646A

; Patent No. 5877001

; GENERAL INFORMATION:

; APPLICANT: Murphy et al.

; TITLE OF INVENTION: Amidases

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

; STREET: 6 BECKER FARM ROAD

; CITY: ROSELAND

; STATE: NEW JERSEY

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH DISKETTE

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: WORD PERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/664,646A

; FILING DATE: June 17, 1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Charles J. Herron

; REGISTRATION NUMBER: 28,019

; REFERENCE/DOCKET NUMBER: 331400-53

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700

; TELEFAX: 201-994-1744

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1869 NUCLEOTIDES

; TYPE: NUCLEIC ACID

; STRANDEDNESS: SINGLE

; TOPOLOGY: LINEAR

; MOLECULE TYPE: DNA

; US-08-664-646A-1

Query Match 7.6%; Score 149.8; DB 2; Length 1869;

Best Local Similarity 52.3%; Pred. No. 3.1e-40;

Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACITATTCTTAACATACACGGGCTCCGCATATGATGACGACATACATATTTTTCAT 1350

Db 1135 CCGGTGATAGTCTTCCTCCACGGCGGCCGAAGGGCATGTACGACACCCGCTTGTCTATC 1194

QY 1351 GAGTTTCAGTGCTGGCGGCAAGGATACCGGTTCGTTTATATCAATCCGAGGAAGC 1410

Db 1195 GAGATCAGCTGATGCGAGCAAGGGCTACTACTGCTCTCGTGAACCCCGCGGCGAC 1254

QY 1411 CACGGCTACGGGACGAAATTTGTAATGCGGTGACAGGAGATTTATGGGAAAGGATAT 1470

Db 1255 GACGGCTATAGCAAGAACCTTCGCGCTCCGCTCTCTGGAGAGGACTGGCTTGGAGACT 1314

QY 1471 GACGATGTGATCAGGCTGTGGATGAGGCTATCAACGAGATCCGCATATTTGATCCTTAAG 1530

Db 1315 GAGGACATATGACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCAGGCGGACAGGAG 1374

QY 1531 CGGCTCGGTGTACGGCGGGAAGCTACGGAGGTTTTATGACCAACTGGATCGTGGGCGAG 1590

Db 1375 CGGCTTGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCGCTTGACTCAG 1434

QY 1591 ACCAACCGCTTTAAAGCTGCCGTACCCAGCGCTGATATCAAAATTTGGATCAGCTTTTCAC 1650

Db 1435 AGCGACTCTTCAAGGACGAGAAATGACGAGAACGGCATAGCTACTTGCTCACCAGCTAC 1494

QY 1651 GGCGTCAGTGATATCGGCTATTTCTTTACAGACTGGCAGCTTGAGCATGACATGTTTGAG 1710

Db 1495 GCCTTCGACATAGGGCTCTGTACGAGCTCGAGGTTCATCGGGCCAAATCCGTTAGAG 1554

QY 1711 GACACAAAGCTCTGGGACCGGTCTTCCTTTAAATACGCAACGAAAGCTGGAGACACCG 1770

Db 1555 AAC--GAGAACTTCAGGAAAGCTCAGCCCGCTGTCTACGCTCAGAACGCTGAAGCGCG 1611

QY 1771 CTTTGTGATCTGATGCGAGCGGATGACCGGATGCCGATGCCGATCGAGCAGCGGAGCAGCTG 1830

Db 1612 ATACTCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACAGAGCTTATGTTTC 1671

QY 1831 TTTATCGCTCTGAAAAAATGGGCAAGAAACCAAGCTTGTCCGTTTTCGGAATGCATCG 1890

Db 1672 TACAAGCTGTCAAGGACATGGGCAAGAGCCTACATAGGATATTTCAAGCGCGCGCC 1731

QY 1891 CACAATTTATCAGCACCAGGACACCCCAAGCAGCGGATCAAGCGGCTCAATATATCAGC 1950

Db 1732 CACGGCCACAGCTCCCGGAAGCCGAGGACACAGCGGCGGAGGCTACAGGCTCTTCATA 1791

QY 1951 TCATGGTTTTCATCAACATCTC 1971

Db 1792 GAGTTCTTCGAGCGCAAGCTC 1812

RESULT 3

US-09-066-285-1

; Sequence 1, Application US/09066285

; Patent No. 5985646

; GENERAL INFORMATION:

; APPLICANT: Murphy et al.

; TITLE OF INVENTION: Amidases

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSLAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,285
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/664,646
FILING DATE: June 17, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herion
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-53
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1869 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-09-066-285-1

Query Match 7.6%; Score 149.8; DB 2; Length 1869;

[illegible][illegible]

```

RESULT 4
US-09-261-006-1
; Sequence 1, Application US/09261006
; Patent No. 6004796
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amidasex
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, (
; ADDRESSEE: CECCHI, STEWART & OLST
; STREET: 6 BECKER FARM ROAD
; CITY: ROSLAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/261,006
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/664,646
; FILING DATE: June 17, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; US-09-261-006-1

```

```

Query Match          7.6%; Score 149.8; DB 3; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

```

[illegible]

```
1471 GAGGATGTGATGAGGCTGTGGATGAGGCTATCAACGAGATCGCATATATGATCCTAAG 1530
1315 GAGGACATATGAACGCATCGAGGAGTTCTTCAAGCTCGAACCGAGCGGAGGAG 1374
1531 CGGCTCGGTGATCAGCGGCGGAAGCTACGAGGTTTATGACCAACTGGATCGTCGGCAG 1590
1375 CGGTTGGAAATAAGCGGCAATAAGCTACGCGGCTTCATGACCAACTGGGCTTGACTCAG 1434
1591 ACGAAGCGCTTAAAGCTGCGCTTACCAGGCTCGATATCAAAATGGATGATGATCCTAC 1650
1435 AGGACGCTCTCAAGGAGGAATAAGCGAGAAAGGCAATAAGCTACTGGCTCACCAGCTAC 1494
1651 GCGCTCAGTCATATCGGCTATTTCTTACAGACTGGCAGCTTGAGCATGACATGTTTAC 1710
1495 GCCTTCTCGGACATAGGCTCTGTTAGGAGTACGAGCTCGAGGCTCGGCGCAAAATCCGCTTAGAG 1554
1711 GACACGAAAGCTCTGGGACCGCTCTCTTTAAATACGAGCAAAACGTCGAGACACCG 1770
1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTCTACGCTCAGAGCTGAAGCGCGG 1611
1771 CTTTGTATCTGCATGCGGAGCGGATGACCGATGCGCGCTGTCTACGCTCAGAGCGGAGCGG 1830
1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCGCTCGACGAGGCTTATGTTTC 1671
1831 TTTATCGCTCTGAAATAAATGGCAAGCAAAAGCTTGTCCGTTTCCGAATGCAATCG 1890
1732 CAGGCGCACAGCTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1791
1951 TCATGTTTGTATCAACATCTC 1971
1792 GAGTTCTTCGAGCGCAAGCTC 1812

RESULT 5
US-09-51-088-1
; Sequence 1, Application US/08951088
; Patent No. 6136583
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amides
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/664,646
; FILING DATE: June 17, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
US-08-951-088-1

Query Match 7.6%; Score 149.8; DB 3; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATTCTTAAACATACACGCGGTCCGCATATGATGATGACGACATATATTTTCAT 1350
Db 1135 CCGGTGATAGTTCTTCCACGCGGCGGAGGAGGATGATGACGACACCGCTTGGCTGTCTAC 1194
QY 1351 GAGTTTCAGTGTGCGGCGGAAAGATACGCGGTCTGTTTATATCAATCCGAGAGGAGC 1410
Db 1195 GAGATGACGCTGATGCGGAGCAAGGCTACTACTGCTGCTGCTGTAACCCGCGGCGAGC 1254
QY 1411 CACGGCTACGCGGAGCAAAATTTGTAATGCGGTCTCAGAGGAGATTTATGGGGAAGATAT 1470
Db 1255 GACGGCTATAGCAAGACTTCGCGCTCCGCGTCTCTGAGAGGAGTGGCTTGGAGGACTTT 1314
QY 1471 GACGATGTGATGAGGCTGTGGATGAGGCTATCAACGAGATCGGATATGATTCCTTAAG 1530
Db 1315 GAGGACATAATGAACGCGCATCGAGGAGTTCTTCAAGCTCGAACCGGAGCGGAGGAG 1374
QY 1531 CGGCTCGGTGTCACGCGGCGGAGCTACGAGGTTTATGATCAAACTGGATCGTCGGGCG 1590
Db 1375 CGGTTGGAAATAAGCGGCAATAAGCTACGCGGCTTCATGACCAACTGGGCTTGGACTCAG 1434
QY 1591 ACGAAGCGCTTAAAGCTGCGCTTACCAGGCTCGATATCAAAATGGATGATGATCCTAC 1650
Db 1435 AGGACGCTCTTCAAGGAGGAATAAGCGAGAAAGGCAATAAGCTACTGGCTCACCAGCTAC 1494
QY 1651 GCGCTCAGTCATGCGGCTATTTCTTACAGACTGGCAGCTTGAGCATGACATGTTTAC 1710
Db 1495 GCCTTCTCGGACATAGGCTCTGTTAGGAGTACGAGCTCGAGGCTCGGCGCAAAATCCGCTTAGAG 1554
QY 1711 GACACGAAAGCTCTGGGACCGCTCTCTTTAAATACGAGCAAAACGTCGAGACACCG 1770
Db 1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTCTACGCTCAGAGCTGAAGCGCGG 1611
QY 1771 CTTTGTATCTGCATGCGGAGCGGATGACCGATGCGCGCTGTCTACGCTCAGAGCGGAGCGG 1830
Db 1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCGCTCGACGAGGCTTATGTTTC 1671
QY 1831 TTTATCGCTCTGAAATAAATGGCAAGCAAAAGCTTGTCCGTTTCCGAATGCAATCG 1890
Db 1672 TACAAGCTGCTCAAGGAGCATGGGCAAGGAGGCTTACATAGCGATATTCAGGCGGCGCC 1731
QY 1891 CACAATTTATCAGCACCGCACCCAGCAAGGAGCGGATCAAGCGGCTGAATATATACAGC 1950
Db 1732 CAGGCGCACAGCTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1791
QY 1951 TCATGTTTGTATCAACATCTC 1971
Db 1792 GAGTTCTTCGAGCGCAAGCTC 1812

RESULT 6
US-09-609-566-1
; Sequence 1, Application US/09609566
; Patent No. 6429004
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amides
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CBCCHI, STEWART & OLSTEIN
```

```

; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/609,566
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/427,372
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; US-09-609-566-1

Query Match 7.6%; Score 149.8; DB 4; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATTCTTAACATACACGGCGGTCGGCATATGATGTCAGGCACATACATATTTTCAT 1350
DB 1135 CCGGTGATAGTCTTCGTCACGGCGGGCGGAGGATGTCAGGCACCGGCTTCGTCTAC 1194

QY 1351 GAGTTTCAGGTGTCGGGGGAAAGGATACGGCGTCTGTTTATATCAATCCGAGAGGAGC 1410
DB 1195 GAGATCAGGTGATGCGGAGCAAGGCTACTACTGCTCTGCTGAAACCCGCGGCGCAGC 1254

QY 1411 CACGCTACGGGAGGATTTGTAATCCGCTCAGAGGAGATTATGGGGAAGGATTAT 1470
DB 1255 GACGCTATAGCGAGACTTCGCGCTCCGCTCTCGAGAGGACTGGCTTGGAGGACTTT 1314

QY 1471 GACGATGTGATGAGGCTGTGGATGAGCTATCAAGAGGATCGGCATATTTGATCCTAAG 1530
DB 1315 GAGGACATATGACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCGGAGGAG 1374

QY 1531 CGGTCGCTGTCACGGCGGAGGATCAGGAGGTTTATGACCAACTGGATCGTGGGAG 1590
DB 1375 CGCGTTGGAATAACGGGCTAAGCTACGGCGGCTTCATGACCAACTGGGCGCTTACTAC 1434

QY 1591 ACGAACCGCTTTAAAGCTGCGTTACCGAGGCTCGATATCAAAATTTGATCAGCTTTCAC 1650
DB 1435 AGCGACCTTTCAAGCAGGAGTAAGCGAGAACGCGCATAGCTACTGGCTCACCAGCTAC 1494

QY 1651 GCGCTCAGTATATCGGCTATTCTTTACAGACTGGCAGCTTGAGCATGACATGTTTGA 1710
DB 1495 GCCTTCGCGACATAGGCTCTGGTACGACGCTCGAGGTCATCGGGCAAAATCCGTTAGAG 1554

QY 1711 GACACAGAAAGCTCTGGGACCGGTCCTCTTTAAATACGAGCAAAAGCTGGAGACAGC 1770
DB 1555 AAC---GAGAATTCAGGAAGCTCAGCCCGCTGTCTACGCTCAGAGCTGAAGGCGCGC 1611

QY 1771 CTTTGTACTGATGGGAGCGGATGACCGGATCCCGATCGGAGCGGAGGAGGAGCTG 1830
DB 1612 ATACTCCTTAATCCACTCGCTTGAGGAGACTACCGCTGTCCGCTCGACGAGGAGCTTATGTC 1671

; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/609,566
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/427,372
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; US-09-609-566-1

Query Match 7.6%; Score 149.8; DB 4; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATTCTTAACATACACGGCGGTCGGCATATGATGTCAGGCACATACATATTTTCAT 1350
DB 1135 CCGGTGATAGTCTTCGTCACGGCGGGCGGAGGATGTCAGGCACCGGCTTCGTCTAC 1194

QY 1351 GAGTTTCAGGTGTCGGGGGAAAGGATACGGCGTCTGTTTATATCAATCCGAGAGGAGC 1410
DB 1195 GAGATCAGGTGATGCGGAGCAAGGCTACTACTGCTCTGCTGAAACCCGCGGCGCAGC 1254

QY 1411 CACGCTACGGGAGGATTTGTAATCCGCTCAGAGGAGATTATGGGGAAGGATTAT 1470
DB 1255 GACGCTATAGCGAGACTTCGCGCTCCGCTCTCGAGAGGACTGGCTTGGAGGACTTT 1314

QY 1471 GACGATGTGATGAGGCTGTGGATGAGCTATCAAGAGGATCGGCATATTTGATCCTAAG 1530
DB 1315 GAGGACATATGACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCGGAGGAG 1374

QY 1531 CGGTCGCTGTCACGGCGGAGGATCAGGAGGTTTATGACCAACTGGATCGTGGGAG 1590
DB 1375 CGCGTTGGAATAACGGGCTAAGCTACGGCGGCTTCATGACCAACTGGGCGCTTACTAC 1434

QY 1591 ACGAACCGCTTTAAAGCTGCGTTACCGAGGCTCGATATCAAAATTTGATCAGCTTTCAC 1650
DB 1435 AGCGACCTTTCAAGCAGGAGTAAGCGAGAACGCGCATAGCTACTGGCTCACCAGCTAC 1494

QY 1651 GCGCTCAGTATATCGGCTATTCTTTACAGACTGGCAGCTTGAGCATGACATGTTTGA 1710
DB 1495 GCCTTCGCGACATAGGCTCTGGTACGACGCTCGAGGTCATCGGGCAAAATCCGTTAGAG 1554

QY 1711 GACACAGAAAGCTCTGGGACCGGTCCTCTTTAAATACGAGCAAAAGCTGGAGACAGC 1770
DB 1555 AAC---GAGAATTCAGGAAGCTCAGCCCGCTGTCTACGCTCAGAGCTGAAGGCGCGC 1611

QY 1771 CTTTGTACTGATGGGAGCGGATGACCGGATCCCGATCGGAGCGGAGGAGGAGCTG 1830
DB 1612 ATACTCCTTAATCCACTCGCTTGAGGAGACTACCGCTGTCCGCTCGACGAGGAGCTTATGTC 1671

; SEQUENCE 1, Application US/09609570
; Patent No. 6465204
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amidases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/609,570
; FILING DATE: 30-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/427,372
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-609-570-1

Query Match 7.6%; Score 149.8; DB 4; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATTCTTAACATACACGGCGGTCGGCATATGATGTCAGGCACATACATATTTTCAT 1350
DB 1135 CCGGTGATAGTCTTCGTCACGGCGGGCGGAGGATGTCAGGCACCGGCTTCGTCTAC 1194

QY 1351 GAGTTTCAGGTGTCGGCGGCAAGGATACGGCGTCTGTTTATATCAATCCGAGAGGAGC 1410
DB 1195 GAGATCAGGTGATGCGGAGCAAGGCTACTACTGCTCTGCTGAAACCCGCGGCGCAGC 1254

QY 1411 CACGCTACGGGAGGATTTGTAATCCGCTCAGAGGAGATTATGGGGAAGGATTAT 1470
DB 1255 GACGCTATAGCGAGACTTCGCGCTCCGCTCTCGAGAGGACTGGCTTGGAGGACTTT 1314
```


; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.8%; Score 35.6; DB 4; Length 289;

Best Local Similarity 5.2%; Pred. No. 0.058; Mismatches 108; Indels 0; Gaps 0;

Matches 11; Conservative 93; Mismatches 108; Indels 0; Gaps 0;

QY 299 AAGCAAGAACTGACTGATATCCCATATGCGGTGTCAAAGCGCTATGGTCCCGGAGC 358

Db 45 RARCRARURGRNR 104

QY 359 GTGAATCGATTCTGCTACTATCATGTTGGGAGAGGGGAAAGCAATTGATGACCGAGAA 418

Db 105 RNR 164

QY 419 AAACAGACGAGCAGCTATCAACTGTTGAAGTCGAAGGCTCTCTACAAACGGGAGC 478

Db 165 RNR 224

QY 479 GCAAAGGCTGACGAGAGGTGCGTATGCCAG 510

Db 225 RCRURGRGRURARARCRURCRURGRCRG 256

RESULT 12

US-09-244-796-17

; Sequence 17, Application US/09244796

; Patent No. 6281344

; GENERAL INFORMATION:

; APPLICANT: Szostak, Jack W.

; APPLICANT: Roberts, Richard W.

; APPLICANT: Liu, Rihe

; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN

; FILE REFERENCE: 00786/350007

; CURRENT APPLICATION NUMBER: US/09/244,796

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/035,963

; EARLIER FILING DATE: 1997-01-27

; EARLIER APPLICATION NUMBER: 60/064,491

; EARLIER FILING DATE: 1997-11-06

; EARLIER APPLICATION NUMBER: 09/007,005

; EARLIER FILING DATE: 1998-01-14

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 17

; LENGTH: 289

; TYPE: RNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Translation template

; NAME/KEY: misc_feature

; LOCATION: (1)...(289)

; OTHER INFORMATION: n = A,T,C or G

US-09-244-796-17

Query Match 1.8%; Score 35.6; DB 4; Length 289;

Best Local Similarity 5.2%; Pred. No. 0.058; Mismatches 108; Indels 0; Gaps 0;

Matches 11; Conservative 93; Mismatches 108; Indels 0; Gaps 0;

QY 299 AAGCAAGAACTGACTGATATCCCATATGCGGTGTCAAAGCGCTATGGTCCCGGAGC 358

Db 45 RARCRARURGRNR 104

QY 359 GTGAATCGATTCTGCTACTATCATGTTGGGAGAGGGGAAAGCAATTGATGACCGAGAA 418

Db 105 RNR 164

QY 419 AAACAGACGAGCAGCTATCAACTGTTGAAGTCGAAGGCTCTCTACAAACGGGAGC 478

Db 165 RNR 224

QY 479 GCAAAGGCTGACGAGAGGTGCGTATGCCAG 510

Db 225 RCRURGRGRURARARCRURCRURGRCRG 256

RESULT 13

US-08-512-955-3/C

; Sequence 3, Application US/08512955

; Patent No. 5976536

; GENERAL INFORMATION:

; APPLICANT: Stephens, David S.

; APPLICANT: Kahler, Charlene M.

; TITLE OF INVENTION: Neisseria Mutants, Lipooligosaccharides

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee and Winner, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: US

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; APPLICATION NUMBER: US/08/512,955

; FILING DATE: 09-AUG-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Ferber, Donna M.

; REGISTRATION NUMBER: 33,878

; REFERENCE/DOCKET NUMBER: 12-95

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 499-8080

; TELEFAX: (303) 499-8089

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 729 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ANTI-SENSE: NO

; ORGANISM: Neisseria meningitidis

; STRAIN: NMB

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..729

US-08-512-955-3

Query Match 1.7%; Score 34.2; DB 2; Length 729;

Best Local Similarity 58.3%; Pred. No. 0.34;

Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1418 ACGGCGAGGAATTTGTAATCGGTCAGAGAGATATGGGGAAGGATTATGACGATG 1477

Db 206 ACGGCGGCGGTCGCGGGATGTTTCGCGCGCATGATGCGGTATTTGAGCCCGACGATG 147

QY 1478 TGATCGAGCTGTGGATGAGCTATCAACGAGATCCGCATAT 1520

Db 146 TGTTCGCGAGAGGTGAGGATTTGACCCAGACCGCGCAT 104

RESULT 14

US-09-221-017B-661

; Sequence 661, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 661:
SEQUENCE CHARACTERISTICS:
LENGTH: 3274 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1...3274
US-09-221-017B-661

Query Match 1.7%; Score 33.6; DB 4; Length 3274;
Best Local Similarity 56.2%; Pred. No. 1.6;
Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 386 TGGGAGAGGGGAAAGCAATTGATGACCGGAGAAAAACAGACGACGACGCTATGAACTG 445
DB 419 TGGCAGACGAAAAAGATCGGATCAATCAGATTGACGAGCTCAAGATAGATAGAGGAG 478
QY 446 TTGAAGTCAGGGCTCTCCACAAAACGGGACGGCAAGGGCTGACGAGAG 497
DB 479 CTGAAGTTTCAGCGCGATCGTCCGACGAGGGGCGGATACGACGAGTGG 530

RESULT 15

US-08-557-139-1/c
; Sequence 1, Application US/08557139
; Patent No. 5827730
; GENERAL INFORMATION:

APPLICANT: Pedersen, Oluf
APPLICANT: Bjorbak, Christian
APPLICANT: Frederiksen, Kathrine A.
TITLE OF INVENTION: MUTANT DNA ENCODING INSULIN RECEPTOR
TITLE OF INVENTION: SUBSTRATE 1
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58277300 No. 5827730disk of No. 5827730th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,139
FILING DATE: 12-FEB-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4041.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6152 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 581...4309
US-08-557-139-1
Query Match 1.7%; Score 33.6; DB 1; Length 6152;
Best Local Similarity 59.4%; Pred. No. 2.5;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 426 GCAGGACAGCTATGAACCTGTGGAAGTGCAGGCTCTCTCTACAAACGGGACGCAAGG 485
DB 914 GTAGGAGAGCCCTGGTACCAGCTGTCTTGTCTGGCCCTCGCTGTCCGCCGCGATGGCAAAGT 855
QY 486 GCTGACGAGAGGTGCGTATGCCAGCTTGTGCTTGT 521
DB 854 GCTCTCCGGGTGTAGAGAGCCACCAGGTGCTTGT 819
Search completed: July 4, 2003, 09:58:25
Job time : 124 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 09:28:15 ; Search time 355 Seconds
(without alignments)
8666.800 Million cell updates/sec

Title: US-09-462-845-1

Perfect score: 1971

Sequence: 1 atgaaaagctgataaccgc.....catggtttgatcaaatctc 1971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUBseq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUBseq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUBseq.*
4: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUBseq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUBseq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMBseq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUBseq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMBseq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUBseq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMBseq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUBseq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMBseq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUBseq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMBseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	807.8	41.0	1958	10 US-09-974-300-224	Sequence 224, App
2	192.2	9.8	477	10 US-09-974-300-4866	Sequence 4866, App
3	149.8	7.6	1869	10 US-09-966-803-1	Sequence 1, Appli
4	82.4	4.2	585	10 US-09-974-300-4838	Sequence 4838, App
5	46.4	2.4	1797	10 US-09-974-300-338	Sequence 338, App
6	37	1.9	567	9 US-10-123-155-254	Sequence 254, App
7	36.8	1.9	400	9 US-10-184-644-530	Sequence 530, App
8	36.8	1.9	400	9 US-10-184-634-530	Sequence 530, App
9	36.4	1.8	342	10 US-09-960-352-4863	Sequence 4863, App
10	36.4	1.8	1839	9 US-10-156-761-4548	Sequence 4548, App
11	36.4	1.8	9025608	9 US-10-156-761-1	Sequence 1, Appli
12	36	1.8	568	10 US-09-974-300-2956	Sequence 2956, App
13	35.4	1.8	380	10 US-09-983-965-3736	Sequence 3736, App
14	35.4	1.8	3316	9 US-10-242-056-58	Sequence 58, Appli
15	34.8	1.8	2374	9 US-10-237-271-6	Sequence 6, Appli
16	34.8	1.8	2383	12 US-10-044-090-521	Sequence 521, App
17	34.6	1.8	1141	9 US-10-184-644-120	Sequence 120, App
18	34.6	1.8	1141	9 US-10-184-634-120	Sequence 120, App
19	34.2	1.7	15150	9 US-10-146-473-9	Sequence 9, Appli

20	34.2	1.7	15164	9 US-10-097-534-19	Sequence 19, Appli
21	34.2	1.7	15164	10 US-09-919-497-20	Sequence 20, Appli
22	33.6	1.7	444	9 US-10-123-155-498	Sequence 498, App
C 23	33.6	1.7	541	9 US-10-184-644-152	Sequence 152, App
C 24	33.6	1.7	541	9 US-10-184-634-152	Sequence 152, App
C 25	33.6	1.7	5828	9 US-09-436-184-6	Sequence 6, Appli
C 26	33.6	1.7	5828	10 US-09-903-248-6	Sequence 6, Appli
C 27	33.6	1.7	5828	10 US-09-859-604-6	Sequence 6, Appli
C 28	33.6	1.7	5828	10 US-09-903-063-6	Sequence 6, Appli
C 29	33.6	1.7	5828	10 US-09-903-216-6	Sequence 6, Appli
C 30	33.6	1.7	5828	10 US-09-903-199-6	Sequence 6, Appli
C 31	33.6	1.7	5828	10 US-09-880-107-3021	Sequence 3021, App
C 32	33.6	1.7	5828	10 US-09-903-023-6	Sequence 3, Appli
C 33	33.6	1.7	20300	9 US-09-939-209A-3	Sequence 6, Appli
34	33.2	1.7	837	9 US-10-184-644-352	Sequence 352, App
35	33.2	1.7	837	9 US-10-184-634-352	Sequence 352, App
36	33	1.7	567	9 US-10-123-155-476	Sequence 476, App
37	33	1.7	671	9 US-10-184-644-346	Sequence 346, App
38	33	1.7	671	9 US-10-184-634-346	Sequence 346, App
39	33	1.7	4108	10 US-09-925-301-493	Sequence 493, App
40	32.8	1.7	888	10 US-09-770-445-502	Sequence 502, App
41	32.4	1.6	504	9 US-09-918-995-9290	Sequence 9290, App
42	32.4	1.6	7680	10 US-09-939-581A-3	Sequence 3, Appli
43	32.2	1.6	514	9 US-09-918-995-32645	Sequence 32645, A
44	32.2	1.6	2706	10 US-09-815-242-5937	Sequence 5937, A
C 45	32	1.6	438	9 US-09-931-457A-27	Sequence 27, Appli

ALIGNMENTS

RESULT 1

US-09-974-300-224
; Sequence 224, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 224
; LENGTH: 1958
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-224

Query Match	41.0%;	Score	807.8;	DB ID:	Length	1958;
Best Local Similarity	64.3%;	Pred. No.	6.5e-258;			
Matches 1268;	Conservative	0;	Mismatches	687;	Indels	18;
Gaps	3;					
QY	1	ATGAAAAAGCTGATAACCGCAGACGACATCATCAGCGATGTCTCTGTGACCGATCTCTCAA	60			
DB	1	ATGAAGCAGCTGATAACCGAAAAAGACCTCATCAAGCTGTATCCATTACCATCCCGAG	60			
QY	61	TACGCCCGCAGCGGTACCCGTCGCCCATATGTAATAATCACAAGTAAATCAAGAGAAAGAT	120			
DB	61	TATTCGCCGATGTGTGAAAAATCGCTACGTCCAAACAAAGTAAGAAAAACAGGAC	120			
QY	121	TCGTATACATCAATATATATGATCTATGAAACGAAACGGGAGGATCTGCTCTTGGACA	180			
DB	121	TCATATGATTCGATATCATGATCTACGACCGGAAAAACAAAGCTTCGGTCAATGGAGC	180			
QY	181	CATGAGAAAAAGCGAACGACCGACCCCAAGATGGTCTCCGACGGCGCGACGCTTCCTTT	240			

181 TTTGGGAAGGCAAGAACACAGCATCCCGCTGCTGCGCGAGCGCAATACCTTGCATTT 240
QY
241 ATTTCGATCGAAGCGGATGCGGACAGCTTTATATCATGATGAGCACTGAAGCGGAGAA 300
Db
241 ACCTCAATCGCGA--GGAAACCGCACAAATTTACGTGATAGCGCGCGCGGGGAGAA 297
QY
301 GCAGGAAGAACTGACTGATATCCCATATGCGTGTCAAGCGGCTATGTCCTCCCGGAGCGGT 360
Db
298 GCAGGAAGAACTGACTGATATCCCATATGCGTGTCAAGCGGCTATGTCCTCCCGGAGCGG 357
QY
361 GAATCGATTTCTGCTCATATCATGTTTGGGAGGCGGGAAGCAATGATGACCGGAGAAAA 420
Db
358 AAGTCATTTGTTTTCAGTGAAGCTGACAAAAGAGAAAGCGTGCAGAGATGAGAAGAA 417
QY
421 ACAGAGAGAGACAGCTATGAACCTGTTGAAGTGAAGCGCTCTCTCAACACGCGGAGCGG 480
Db
418 ACCGAGATTTGAAGACCAACGACCGCTGGAAGTCGATTTCTTTGTCTATATAAGCAGAGCGG 477
QY
481 AAAGGGCTGACGAGAGTGGTGTATGCCAGCTTCTGCTGTGTCAGCGTAAAGTCCGGTGAG 540
Db
478 CAAGGGTTTAAAGGGGAAAGTATACCGAGCTTCTCTGTTCAAGTGGAAACCGGGAA 537
QY
541 ATGAAGAGCTGCAAGTCAAGCTCAAGCTGATCATGGTGTATGCTGTTTTCCTGACGGC 600
Db
538 ATGAAGAGCTGCAAGTCAAGCTCAAGCTGATCATGGTGTATGCTGTTTTCCTGACGGC 597
QY
601 AATGGCTGTTTCTCAGCTAATTTAACTGAACAGATGATGCCAGCAACCGCATGAT 660
Db
598 GATCAGATTCATTTTTCGCGCAATCAACAGGATATGCGTGT-----GAACGAT 645
QY
661 GTTTACATAATGTCACCTGGAGTCTGAGATCTTAAAGCAGGTTTACACCTCATCGCGGTCA 720
Db
646 GTCTACTTGTGAACCTTTTCGACAGGTTGATTTGAAGCGCTTCACAGGTCAAAACGGCAT 705
QY
721 TTCGATCAAGCTCATTTTTCACAGAGAGGATGATCTGCTTTTCTGTTGGAATGAAGAA 780
Db
706 TTTTCTTCACTATGTTTCTCTGAGCGCAACACCTTGCATTTATCGGAAACGAGAAA 765
QY
781 GAATATAAGATGTCACGCTCTCAAGCGTGGCTGTATGATATCGAACAAGCGCGCTC 840
Db
766 GAATATCAAAACCGCAGCTCGATAAGCATGCTGTATGATATTAACACAGGAAACTA 825
QY
841 ACATGCTTTACTGAGATGCTGGAGCTTCAATTTAGCGGATGCGCTGATTTGAGATTCATG 900
Db
826 ACCTGCTGACGGAATGCTTGACGCTCATCTACGCGATGCGGTGGCAGGAGACAGCCTT 885
QY
901 ATCGGTGCTGAACAGCGCCGATTTGGCAAGGACAGCAAGGCTTTATGTCATC 960
Db
886 GTCGGAGCGGTTCTTCGAAACCGCGCTGGACCAAGACGGAACGCGCTTTATGTCATC 945
QY
961 GGCACAGATCAAGGAGTACGCGCATCTATTTATTTGATTAAGGCGCTTGTGTATCCG 1020
Db
946 GGATCTCACCAGGTTCTACAGGCTTTACTACATATCAATCAAGAGGCTTGTATCCG 1005
QY
1021 ATTGCTCGGAAAGAGTACATCAATGCTTTTCTTCTTCACTGATGAACAGCACTT 1080
Db
1006 GTCGGTTGGAAAGAGCATGTGAACGATTTACGCTGATCCGCGAGAAAGTGGCTTT 1065
QY
1081 ATTCCAGTGTGCAAAAGCGGACAGCGGAGTGTGCTTTACAGTATCC--CGTGTGA 1137
Db
1066 GCGGATCCATAGCCCTGCGCGTGGCGGAGCGGAGCTTTACCTCAATTTGGGTGAA 1125
QY
1138 CAGGAAGAGAAACAGCTGACTGCGGCGGAATGACAAGTTTGTGAGGAGCATACATATCA 1197
Db
1126 GAAAGCGGAGCGACTAACAGATGCCAACCAAGGTTTACGGAAGACATATCATATCG 1185
QY
1198 ATACCTGAAGAGATTCATATGCTTACAGAACGCGCTGATGGTGAACGCGTGTATG 1257
Db
1186 GAACCGGAGGAGCTTCAATTTTACGCGAGCGGAGCGGCTGTGACAATTTACCGGATGGCTGATA 1245
QY
1258 AGGCTCTCACAAATGGAGGTGAGACAACATATCCACTTATCTTAAACATACACGCGGT 1317
Db
1246 AAGCCCGCCCAATATGAAAGAGAAACACGATATCGGCTCATCTCGAAGTGCACGCGGA 1305

QY
1318 CGCATATGATGTCAGGACATACATATTTTTCATGAGTTTCAGGTGCTGGCGGCAAGGA 1377
Db
1306 CTTACGCCATGATGCGAATGCTTATTTTCATGATTTCAATGCTGGCGGCAAGGA 1365
QY
1378 TACGCGGTGCTTTATATCAATCCGAGAGAGCCAGGCTACGGGAGGAATTTGTGAAT 1437
Db
1366 AGCGCGTCTATATGTTTAAACCGCGGGAAGCCAGGCTACGACAGGATTTGTCAAC 1425
QY
1438 GCGGTACAGGAGATATCGGGAAGGATTTATGACGATGATGACGAGCTGTGGATGAG 1497
Db
1426 GCGGTGAGGCGGACTACGCGGTGGAGATTTTAAGGATGATTAAGCGGCTGTGGACAC 1485
QY
1498 GCTATCAACAGAGTCCGCATATTTGATCCTAAGCGGCTCGGTGCTACGGCGGGAAGCTAC 1557
Db
1486 GTCCTGAACATATGATTTGCTGATCAAGAAAGGCTCGGCATCACCGGAGGAAGCTAC 1545
QY
1558 GGAGGTTTTATGACCAACTGGATGCTGGGAGAGCAACCGCTTTAAAGCTGCGGTACC 1617
Db
1546 GCGGCTTTTATGACAACTGGGCTGTCGCGCACACGAAAGCGCTTCAAGGCGCGAGTACA 1605
QY
1618 GAGCGCTCATATCAAAATTTGATGCTGCTTTCACGCGCTCAGTATGATATCGGCTATTTCTTT 1677
Db
1606 CAGAGTCCATTTCCAACTGGATCAGTTTTTACGGGTTAAGCGGATCGGCTATTTTTC 1665
QY
1678 ACAGCTGTCAGCTTGAGCATGACATGTTTGGAGCACACAGAAAGCTTCTGGGACCGGTCT 1737
Db
1666 ACAGCTGTCAGCTGCTGTCAGACCTTTTGAAGATTCAGGCAAACTGTGGGAGCATTCG 1725
QY
1738 CTTTAAATAGCAGCAACAGTGGAGACACCGCTTTTGTACTGTCATGCGGAGCGGAT 1797
Db
1726 CCGCTCAATAGCGGACAAAGTGGAGACCGCTTCTCTCATCTGCGATGGCGAACGGAC 1785
QY
1798 GACCGATGCGGATGAGCGGAGCGGAGCTGTTTATGCTGCTGAAAAAATGGCAAG 1857
Db
1786 GACAGTGGCGGTTGAACAGCGGAGACAGCTGTTTACGCGCTGAAAAAATGGGAAA 1845
QY
1858 GAAACCAAGCTTGTGCTGTTTTTCCGAATGATCGACAAATTTATCACCGCGGACACCA 1917
Db
1846 GAAGTAAAACTGTCAGATTTCCGAAACGCTCACAGATTTATCAAGAGCGGACATCCG 1905
QY
1918 AGACAGCGGATCAAGCGCTGAATTTATATGATGCTCATGCTTGTGATCAACATCT 1970
Db
1906 AAGCAGCGGATCAGCGGCTTGATATATTTGAGGCTGGTTGAGCTTATCT 1958

RESULT 2

US-09-974-300-4866
; Sequence 4866, Application US/059974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4866
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-4866

Query Match 9.8%; Score 192.2; DB 10; Length 477;
Best Local Similarity 65.1%; Pred. No. 4.3e-53;
Matches 300; Conservative 0; Mismatches 158; Indels 3; Gaps 1;


```

; PRIORITY APPLICATION NUMBER: 09/580,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4838
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(585)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4838

```

```

Query Match      4.2%; Score 82.4; DB 10; Length 585;
Best Local Similarity 48.0%; Pred. No. 2e-16;
Matches 233; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

QY 587 TTTCTCCTGACGGCAATGCTTGTCTTCTCAGCTAATTTAACTGGAACAGATGATGCCA 646
Db 1 TTTTACGAATGAGTGTGCTTGTCTTGTAGCCGCGCCATTAAGGAAACAAATGATTTTA 60
QY 647 GCAAGCGCGATGATGTTTACATAATGTCACCTGGAGTCTGGAGATCTTAAAGCAGGTATCAC 706
Db 61 CTTTTCGAGTGTGATGTTTGTGTTTATGCTTGGATNGGGAGCAAAACGCAATACAA 120
QY 707 CTCATCGCGCTCATTCGGATCAAGCTCATTTTCCACGACGGAGGATATCTTCTGCTTTCG 766
Db 121 AGAACCAGATGCTGATCAAAAGCGGATCTGTATCACCGGATGTTCTTATGCTGCTCTCC 180
QY 767 TTGGAATGAAAGGAATATAGATGCTAGCTCTCAAGGCGCTGGCTCTATGATATCG 826
Db 181 TAGGCCATGAGAGAAATACAAACCGGACATGACGAAGCTGTATGTTTATGATGCC 240
QY 827 AACAAAGCCGCTCACATGCTTACTAGATGCTGGAGCTTCATTTAGCGGATGCGCTGA 886
Db 241 AAAACAAGAAATCCGTTTCTTCCGCGCAAGTGGATGCTCAATTTGGCGATGTTGCAA 300
QY 887 TTGGAGATTCATGATCGGTGGTCTGAACAGCGCCGATTTGGACAAGACAGCCCAAG 946
Db 301 TTGGGGAATTTTACAGAAATATGTTTCAAAATGGGTTGAAATGGGAGGAAAGATGTCAGT 360
QY 947 GGTATTTATGTCATCGGCACAGATCAAGGCGATGACGGCATCTATTATATTTTCGATGAAG 1006
Db 361 CCATTATGCGCTGTTTCAAAAGAGCGGATGTAATGTTTGGTCTGTTAGCCCTCGATG 420
QY 1007 GCCTTGTGATPCGATGCTCTGGAAGAGAGTACATCAATAGCTTTTCTCTTTTCACTG 1066
Db 421 GCACAGTCAATCAGGTACGCGATGAAACAAACGTCACATAAATGGATTCGATTTTCAAAATG 480
QY 1067 ATGAA 1071
Db 481 GCCAA 485

```

```

RESULT 5
US-09-974-300-338
; Sequence 338, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27

```

```

; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 338
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-338

```

```

Query Match      2.4%; Score 46.4; DB 10; Length 1797;
Best Local Similarity 44.6%; Pred. No. 0.00042;
Matches 363; Conservative 0; Mismatches 436; Indels 15; Gaps 4;

QY 1094 CAAAGCCGAGACACCGAGTGTGAGCTTTACAGTATCCCGCTTGGACAG---GAAGAGAAAC 1150
Db 902 CAAGCCGAGACCGAGTGTGAGCTTTACAGTATCCCGCTTGGACAG---GAAGAGAAAC 961
QY 1151 AGCTGACTGGCCGGAATGACAAGTTTGTCTAGGAGCATACGATATCAATACCTGCAAGAGA 1210
Db 962 AATTAACGAAACATTCGGTTCGGGTTTCTCCGACACAGGAGCTGTCTTATCTCTGAAGTTG 1021
QY 1211 TTCAATATGCTACAGAAGACGGGTGATGTTGAACGGCTGGCTGTAGAGCCCTGACACAAA 1270
Db 1022 TCACATATCTTCTTTTGTATGGCTGCCAATAGAGGACTGCTGTTTAAAGCCCTCCCGG 1081
QY 1271 TGGAGGTGAGACACATATCCACTTATTTCTTAAACATACACGGGGTCCCATATGATGT 1330
Db 1082 TAGAAG-----CGAACGGCTGGACGATTTATTTGGCCGCGACGGGACCGCAAGATCGAG 1135
QY 1331 ACGGACATACATATTTTTCATGATGTTTCAGGTGCTGGCGGCGAAAGGATACGCGTCTGTT 1390
Db 1136 AAACCTTATGTTCTACGATCTCTTCACTCCCTGCAAAATGGTTTACCAGTTATTCG 1195
QY 1391 ATATCAATCCGAGAGAACCCAGCGCTACGGGCGAGAAATTTGTGAATTCGGTTCAGAGAG 1450
Db 1196 CTCCTAATTCAGGGGATCAGCGAATACGTTTATTCATTTCTTTAAATGGTTGAACAAG 1255
QY 1451 ATTATGGGGAAGGATTTAGCATGTGATGAGCTGTGGATGAGGCTATCAACAGAG 1510
Db 1256 ATTGGGAGACGGCGCGCTCGATATGACCGGGCATCGACTGGCTATTCGATCANA 1315
QY 1511 ATCCGATATTTGATCTTAAGCGCTGCTACGGCGGAGAACTCTTTTAAATGGCGGAGTATGT 1570
Db 1316 A---GCTTCAGACCGCGGAAACTCTTTTAAATGGCGGAGTATGTGAGGATATGT 1372
QY 1571 CCAACTGATGTCGGGCGAGACGACCGCTTTAAAGCTGCGTTTACCAGCGCTCGATAT 1630
Db 1373 CGCTTTTGTCTCCACGGGAGGCATCTCTGAGTATTTCCGCGCGCTGCTGATATATCGGAG 1432
QY 1631 CAAATTTGGATCAGCTTTCAGCGGCTCAGTATGATATCGGCTATTTCTTTACAGACTGGCAGC 1690
Db 1433 TCAGCAACCTGTTTTCGTTCTCAAAACCGTACCTGATTTCTGGCAGCCGATGATGGAA 1492
QY 1691 TTGAGCATGACATGTTTGA---GGACACAGAAAGCTCTGGGACCGGCTCTCTTTTAAAT 1747
Db 1493 AATGGGTGCGCAATCCGAGCGGAGTATGAAAAAATGAAAGCGGATTTCTCTGTTACAT 1552
QY 1748 ACGCAGAAACGTGGAGACACCGCTTTTGATACCTGTCATGCGCGGCGGATCCCGATGCC 1807
Db 1553 ATCTGGAGAAATATGACTCAGCCGATGCTGATCATCCAGGGCGCAACGATCCGCGCTCG 1612
QY 1808 CGATCGACGACGGGAGCAGCTGTTTATCGCTCTGAAAAAATGGGCAAGAAACCAAGC 1867
Db 1613 TAAAGGAGAGTCCGATCAGTCTCGATCAGTTGAGGAAACATGAGGCGGAATATCGAGT 1672
QY 1868 TTCTCCGTTTTCGGAATGCAATGCGACAAATTTATC 1901
Db 1673 ATCTCGTTTGGAAATGAAGGCGACGCTTTTC 1706

```

```

RESULT 6
US-10-123-155-254
; Sequence 254, Application US/10123155
; Publication No. US20030068794A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 254
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-254

Query Match
Best Local Similarity 1.9%; Score 37; DB 9; Length 567;
Matches 36; Conservative 117; Mismatches 238; Indels 0; Gaps 0;

QY 1271 TGGAGGTGACACACATATCCACTTATCTTACATACACGGCGGTCCGATATGATG 1330
Db 166 YGGWNVSMANFRDYANLCFEAFGDRYKHWTTFSDPRMAEKGYETGHHAPGLKRG 225
QY 1331 ACGGACATACATATTTTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1390
Db 226 LYKAHHIIKAHAKTHWISYNTWRSKQGLVIGLSLNCWGPVDSINPKDLBAERYLQ 285
QY 1391 ATATCAATCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1450
Db 286 CLGWFAPIYAGDYPQVMKDYIGRKSABQGLMSRLPVFSLQEKSYIKGTSDFLGLH 345
QY 1451 ATTATGGGGAAGGATATACCATGTGATGATGATGATGATGATGATGATGATGAT 1510
Db 346 TRYITERNYPGRQGSYQNDRLIELVDPNPDLGSKWLYSVPMWFRLLNFAQTQYGD 405
QY 1511 ATCCGCATATTGATCCTAAGCGCTCGGTGTCAGCGGCGGAGGAGGAGGAGGAGG 1570
Db 406 PIYVWENGASQKHFCTQLCDEWRIOYLGKYNEMLKAKDGANIKGYTSWLLDKFE 465
QY 1571 CCACTGATGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1630
Db 466 GYSRDYGFYFFVFEFDRNKRPRKASQVYKKIIANGFPNPNREVESWYLKALET 525
QY 1631 CAAATTGGATCAGCTTTCACGGCGCTCAGTGA 1661
Db 526 MAAEPLLSHMQWTEIVVPTVCSLCVLIITA 556

RESULT 7
US-10-184-644-530
; Sequence 530, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 530
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-530

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 530
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-530

Query Match
Best Local Similarity 1.9%; Score 36.8; DB 9; Length 400;
Matches 22; Conservative 93; Mismatches 132; Indels 0; Gaps 0;

QY 324 ATATGGCGGTGTCACAAAGCGCTATGTCGCCGAGCGGTGAATCGATTCTGCTACTATCAG 383
Db 89 VHIIGALRTSKLLSDPNYGVHLPVAKLRHVRHVMYOWVETEESREYTEDQVKKETRYSYN 148
QY 384 TTTGGGAGAGGGGAAAGCATTTGATGACCGAGGAAAAACAGAGGAGGAGGAGGAGGAGG 443
Db 149 TEWRSEIINSKNFDRIGHKNPSAMAVESFMATAPVQIGREFLSGLIDKVDNFKLSL 208
QY 444 TCTTCAAGTGCACGACGCTCTCTCTACAAAGCGGACGCGGACGAGGAGGAGGAGGAGG 503
Db 209 SKLEDPHVDIIRRGDFYHSENPKYPEVGDLRVSYAGLSGDDPDLGPAHVVTVIAROR 268
QY 504 TGCCGAGCTTGTGCTGTGTCAGCGTAAAGTCGGGTGAGATGAAAGAGCTCACAAGTCACAA 563
Db 269 GDOLYPTSTKSGDTLLHLLHGGDFSABEEVPHRELRSNMTWGLRAAGWMAFMFGLNLMTR 328
QY 564 AGCTGAT 570
Db 329 ILYTLVD 335

RESULT 8
US-10-184-634-530
; Sequence 530, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 530
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-530
```


APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HAITORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 1.8%; Score 36.4; DB 9; Length 9025608;
Mismatches 0; Mismatches 296; Indels 3; Gaps 1;
Matches 227; Conservative 0;

QY 1304 ACATACACGGCGTCCGCATATGATGACGACATACATATTTTCATGATGTTTCAGGTGC 1363
DB 5573387 AATCCACGGCGTCCGCCTGACGACGACGACTCGTTCGGCGGGCGCCGGCCCT 5573328

QY 1364 TGGCGGGAAGGATACGGCGTCTTTATATATCAATCCGAGAGAACCCAGCGTACGGGC 1423
DB 5573327 GCGTGACACGGGTACGGGTCTGCGGGTCAACTACCGGGTGCACCGGATACGGGC 5573268

QY 1424 AGGAATTTGTGATCGGCTCAGAGGAGATTATGGGGGAAGGATTATGACGATGTC 1483
DB 5573267 GGGAGTGGACGACGGCTCAAGCACCGGGTGGTCTGATCGAGCTGGAGGACATCGCG 5573208

QY 1484 AGGCTGTGATGAGGCTATCAACGAGATCCGCATATTCATCTAAGCGGCTCGGTGTC 1543
DB 5573207 CGTGGGAGTGGCGGTGTCGTCGGGCTCGCGACCCGACAGGCTGATCGTACCG 5573148

QY 1544 CGGGCGGAAGCTACGAGGTTTTATGACCAACTGATCGTGGGACAGCAACCGCTTTA 1603
DB 5573147 GCGGTTCTGGGCGGTATCTCACGCTCGGCTCGGACCCAGCGGAGCGTGA 5573088

QY 1604 AAGCTGCGCTTACCGAGCTCGATATCAAAATGGATGATCTTTCAGCGCTCAGTGATA 1663
DB 5573087 CGATCGGATCGCGGCTCGCGTGGCGGCTACGCTACGCGGTACCGAGATGG 5573028

QY 1664 TCGGCTATTTTACAGCTGCGAGCTTGACATGACATGTTTGAAGNCA---CAGAAA 1720
DB 5573027 AGCCGCTGAAGCGGTGACCGGCTGCTCGGGGCGACCGGAGGAGTGGCGGAGC 5572968

QY 1721 AGCTCGGACCGGTCTCTTTTAAATACGACAAACGTTGGAGACACCGCTTTTGATAC 1780
DB 5572967 GGTTCGAGCGCTGCGCGGTGACCTACGTGACGCGGTGAAGCGACCGGTGTACATCT 5572908

QY 1781 TGCATGGCGGATGACCGATCCCGATCGGACGCGGAGCA 1826
DB 5572907 CGGCGGCGCTCAACGATCCGGTGGCGGCTCCGCATCCGCGAGTGAGAA 5572862

RESULT 12

US-09-974-300-2956

Sequence 2956, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION: Methods For Monitoring Multiple Gene
TITLE OF INVENTION: Expression
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2956
LENGTH: 568
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-2956

Query Match
Best Local Similarity 1.8%; Score 36; DB 10; Length 568;
Mismatches 108; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
Matches 108;

QY 392 AGGGGAAAGCATTTGATGACCGAGAAAAACAGACAGGACGATGATGAACCTGTTGAAG 451
DB 317 ATGAAGAAAGGCTCGTCTTCATCATCAGAAAGCTGTTGGATTTTGCACGATGCCGTTT 376

QY 452 TGAAGGCTCTCTCTACAAAGGACCGCAAGGGCTGACGAGAGGTGCGGTATGCCGAGC 511
DB 377 TGAAGGACGCGCAATCATCTGACGATCATCTGACGAGCGGTCGCGGACGCGCC 436

QY 512 TTGCTGTGTGTCAGCGTAAAGTCGGGTGAGATGAAGAGCTGACAAAGTCACAAAGCTGATC 571
DB 437 TCCTCATCTTCTCGATTTCCACGGCGTATTCACAAAGCTGACCGGTATCAAGACTTTC 496

QY 572 ATGCTGATCTGCTTTTCTCTGACGGCAAAATGGCTGTTTCTCTAG 619
DB 497 ATCATTACTGGCGGCTGTTTATGAAATCAAGCGGTTTGATGTAACAG 544

RESULT 13

US-09-983-965-3756/c

Sequence 3756, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Ningbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
FILE REFERENCE: 37-21(10297)C

CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 3756
LENGTH: 380

TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (371)
OTHER INFORMATION:
OTHER INFORMATION: Clone ID: 46-LIB3058-008-Q1-K1-D10

US-09-983-965-3756

Query Match
Best Local Similarity 1.8%; Score 35.4; DB 10; Length 380;
Mismatches 66; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 87 ATATGAAATCAAGTAATCAAGAGAAAGATTCGTATACATCAATATATGATCTA 146

Db 130 AAAATTAATAAACCTAATGATCTTAAGAGATATGAGCATAAATCCGAGAGATGAATCTG 71
QY 147 TGAACCAAAACGGGAGGATCTGTTCTTGACACATGGAGAAAAGCGAAGCACCGA 203
Db 70 GGAAGCAAAACTGTGTCATCTGCTCGATTATATAAGAAATAAATGATCCCGA 14

RESULT 14

US-10-242-056-58
; Sequence 58, Application US/10242056
; Publication No. US20030113323A1
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Orr, Gregory L
; APPLICANT: Merlo, Donald J
; APPLICANT: Roberts, Jean L
; APPLICANT: Rocheleau, Thomas A
; TITLE OF INVENTION: Insecticidal Protein Toxins from
; TITLE OF INVENTION: Photorhabdus
; NUMBER OF SEQUENCES: 88

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dowelanco
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/242,056
; FILING DATE:

; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/743,699
; FILING DATE: 06-NOV-1996

; ATTORNEY/AGENT INFORMATION:
; NAME: Borucki, Andrea T.
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 50301E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-337-4846
; TELEFAX: 317-337-4847

; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4932 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)
US-10-242-056-58

Query Match 1.8%; Score 35.4; DB 9; Length 3316;
Best Local Similarity 13.8%; Pred. No. 2.9;
Matches 46; Conservative 125; Mismatches 162; Indels 0; Gaps 0;

QY 1272 GGAAGGTGACACATATCCACTTATTCTTAACATACACGGCGTCCGCATATGATGTA 1331
Db 2164 RGASNGU(GYTRAAAR)UUAATHRUHSHRVATHRVAVSGYRTYRAA)TRGUGUGUT 2223
QY 1332 CGGACATACATATTTTCATGAGTTTCAGTCTCGCGCGGCGGAAAGATACGGCGTGTGTTA 1391
Db 2224 HRTHRGYTVRASNUY((RRASGY)TRVUAUSTRHASYSASNHAAGYU)ASNYUG 2283
QY 1392 TATCAATCCGAGAGAGAGCCAGCGCTACGGCAGCAATTTGTGAATCGCGTCCAGAGGAGA 1451
Db 2284 USRVAHTRSRRASTRTRTHR)TH((RYSASNHSRYSAAAASNASNARGYSHTYRGN)G 2343
QY 1452 TTATGGGGAAAGGATTATGACGATGTGATCGAGGCTGTGGATGAGGCTATCAACAGAGA 1511
Db 2344 UASNAAGUTHRAAASGYARGASNUHSARGTYR)SRT(HRGNTHRHGYUTHRSRGYAAATH 2403
QY 1512 TCGCATATTGATCCTAAGCGGCTCGGTGTCCAGCGGCGGGAAGCTACGGAGGTTTATGAC 1571
Db 2404 RTYRSRTHR)TYRTHRUSRGUAAASHSRTHRASRYSASNTYR((UGNVACYSUAS 2463
QY 1572 CAACTGGATCGTGGCGAGCAGCAACCGCTTTAA 1604
Db 2464 NVAVATRASHSTYRASGRSR)GYISYSGYAA 2496

RESULT 15

US-10-237-271-6
; Sequence 6, Application US/10237271
; Publication No. US20030096328A1
; GENERAL INFORMATION:
; APPLICANT: THE BURNHAM INSTITUTE
; APPLICANT: SMITH, Jeffrey W.
; APPLICANT: KRIDEL, Steven J.
; APPLICANT: AXELROD, Fumiko T.
; TITLE OF INVENTION: SERINE/THREONINE HYDROLASE PROTEINS AND SCREENING ASSAYS
; FILE REFERENCE: BURN100-1
; CURRENT APPLICATION NUMBER: US/10/237,271
; CURRENT FILING DATE: 2002-09-04
; PRIORITY APPLICATION NUMBER: US 60/317,842
; PRIORITY FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 2374
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-237-271-6

Query Match 1.8%; Score 34.8; DB 9; Length 2374;
Best Local Similarity 49.5%; Pred. No. 3.7;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1721 AGCTCTGGGACCGGTCTCTTTTAAATACGCAAGCAACGTTGGAGACACCGCTTTTGATAC 1780
Db 1968 AGATGCTGGACAATCGCCCATCATATCCCTCAGTGAAGACACCACTGTTACTGA 2027
QY 1781 TGCATGGCGAGCGGATGACCGATCCCGATCGAGCGGAGGAGCTGTTATCGCTC 1840
Db 2028 TGTTGGGCGAGGAGGACCGGCTGTGCGCTTCAAGCAGCGGATGAGTAGTATTACCGTCC 2087
QY 1841 TGAAAAAATGGCAAGCAACCAACCAAGCTTGCTCGTTTCCGATGATCGCACAAATTTAT 1900
Db 2088 TCAAGACCCGGAATGTGCTGTTCGGCTCCTCTCTATCCCAAGCACCCACCGCATTTAT 2147
QY 1901 CA 1902
||

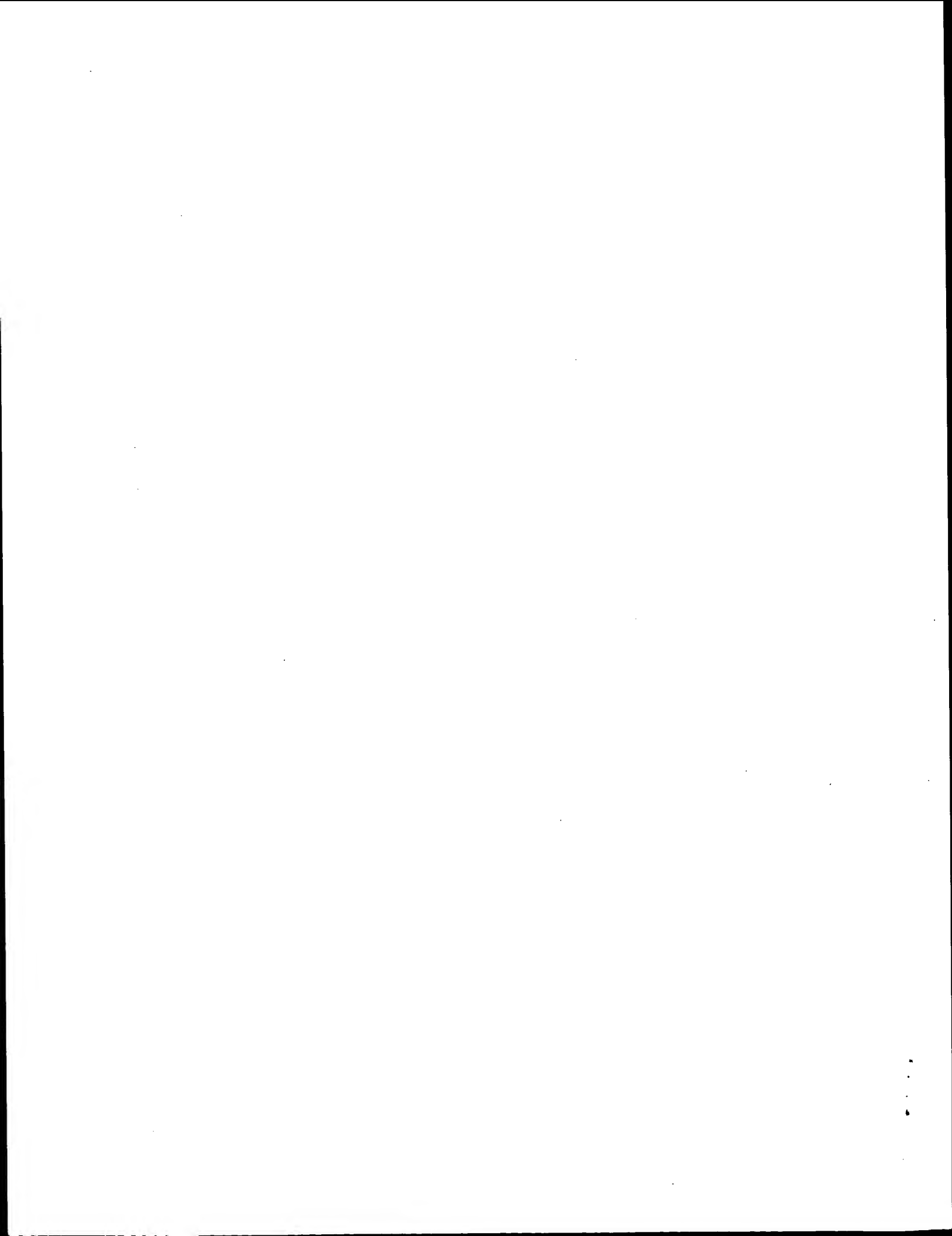
Sat Jul 5 15:57:28 2003

us-09-462-845-1.rnpb

Page 9

Db 2148 CA 2149

Search completed: July 4, 2003, 11:48:54
Job time : 370 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 11:05:37 ; Search time 2559 Seconds

(without alignments)
12474.132 Million cell updates/sec

Title: US-09-462-845-1

Perfect score: 1971

Sequence: 1 atgaaaagctgataaccgc.....catggttgatcaacatctc 1971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estt:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	66	3.3	839	17	AZ535741
C 2	66	3.3	909	17	AZ530780
C 3	50.4	2.6	552	13	BM175145
C 4	48.4	2.5	327	9	AA071581
C 5	46.8	2.4	475	12	BG279712
C 6	43.6	2.2	697	13	BJ153314

C 7	43	2.2	840	13	BI118680
C 8	42.6	2.2	873	17	AZ528669
C 9	42.6	2.2	939	12	BG025910
C 10	42	2.1	695	13	BJ133826
C 11	40.8	2.1	490	13	BJ481303
C 12	40.8	2.1	603	10	AV836520
C 13	40.8	2.1	1101	17	CNS000D1
C 14	39.8	2.0	531	12	BG650557
C 15	39.6	2.0	951	11	AY105542
C 16	39.6	2.0	1449	14	BQ049806
C 17	39.4	2.0	711	10	BE130860
C 18	39.4	2.0	1064	10	BE036593
C 19	37.6	1.9	719	14	BQ443031
C 20	37.4	1.9	1101	17	CNS003EJ
C 21	37.2	1.9	513	10	AW906913
C 22	37.2	1.9	884	17	AZ540694
C 23	37.2	1.9	926	17	AZ689494
C 24	37.2	1.9	932	17	AZ692448
C 25	37.2	1.9	956	17	CNS007BU
C 26	37	1.9	438	13	BM570134
C 27	37	1.9	459	9	AA792627
C 28	37	1.9	491	13	BJ169948
C 29	37	1.9	532	10	AV601962
C 30	37	1.9	579	13	BM565239
C 31	37	1.9	583	13	BM506840
C 32	37	1.9	583	13	BM565238
C 33	37	1.9	588	12	BG800737
C 34	37	1.9	725	13	BI555774
C 35	37	1.9	733	13	BI730952
C 36	37	1.9	747	13	BG971297
C 37	37	1.9	935	14	BQ920566
C 38	37	1.9	939	14	BQ957767
C 39	37	1.9	1000	12	BF534027
C 40	36.8	1.9	377	10	AW353260
C 41	36.8	1.9	406	10	AW353257
C 42	36.6	1.9	717	17	BH028231
C 43	36.4	1.8	418	10	AV590668
C 44	36.4	1.8	457	12	BF707325
C 45	36.4	1.8	526	10	AV605765

ALIGNMENTS

RESULT 1
AZ535741/c
LOCUS
DEFINITION
ACCSSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AZ535741 839 bp DNA linear GSS 03-NOV-2000
ENTCC88TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
AZ535741
AZ535741.1 GI:11092688
GSS.
Entamoeba histolytica.
Entamoeba histolytica.
Fukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 839)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 34
High quality sequence stop: 716.

```

FEATURES
  source
    Location/Qualifiers
      1. 839
      /organism="Entamoeba histolytica"
      /strain="HMI:IMSS"
      /db_xref="taxon:5759"
      /clone_lib="Entamoeba histolytica Sheared DNA"
      /note="Vector: pHOSt; Site 1: Bst I; Constructed at The
      Institute for Genomic Research (TIGR), Rockville, MD.
      Genomic DNA isolated from broth cultures of E. histolytica
      using a method described by Clark and Diamond (Clark,
      C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
      method for isolate identification. Exp. Parasitol.
      77:450.). The DNA was mechanically sheared to give a
      tight size distribution (~2 kb). The v + i method used for
      the library construction is described in detail in Smith,
      H.O. and Venter, J.C. (Making small insert libraries for
      whole genome shotgun sequencing projects. In Genome
      Sequencing: A Practical Approach, eds. M. Vaudin and B.
      Barrell, Oxford University Press, 1999)."
      BASE COUNT      277 a 143 c 107 g 312 t
      ORIGIN

Query Match      3.3%; Score 66; DB 17; Length 839;
Best Local Similarity 48.6%; Pred. NO. 5e-08;
Matches 214; Conservative 0; Mismatches 220; Indels 6; Gaps 1;

Qy 1147 AACAGCTGACTGGCGCAATGACAAAGTTTGTTCAGGGAGCATACGATATCAATACCTGAA 1206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 569 AACAAATCACTCACTAATTAACAAGAGTTCTTCTCAATTAATTTGGAGAATATAAA 510
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1207 GAGATTCAATATGTACAGAGAGCGCGTGATGTTGAACGGCTGGCTGATGAGGCTGCA 1266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 509 GAAATTCATTACACTGGAGCAATTAATGATCAAAATTCATGCAATTTATTACCCACCA 450
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1267 CAATGGAGGTGAGACAACTATCCACTTATCTTACATACACGGCGGCTCCGATATG 1326
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 449 AATATGAACAAACACAAATATCCAGTAATTTTATATACACATGGAGGACCAAGATCA 390
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1327 ATGTACGGACATACATATTTTCAT-----GAGTTTCAGGTGCTGGCGCGAAAGGATAC 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 389 CCATGGACTAATAATTTCCATTATAGATGGAATCCACAGTTTATAGCACACAAAGTTAT 330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1381 GCGTGTGTTTATATCAATCCGAGAGAGAGCCACGGCTAGCGGCGAGGAATTTGTAATGCG 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 ATTGTTTTTGGACCCCAATTTCCATGGATCAGGAAGTTATGGAGATCATTTTAAAGCA 270
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1441 GTCAGAGGATTTATGGGGAAGGATATGACGATGTCATGCAAGCTGTGGATGAGGCT 1500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 269 ATTAGAAGAAATTTGGGAGGATGGCCATTGGAAGATTTAATGAAGGAATGGATTTTA 210
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1501 ATCAACAGAGATCCGCATATTTGATCCTTAAGCGGCTGGGTGCACGGCGGAAAGCTACGGA 1560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 AAAACTAGTGAGCCATTAGTTGATATAGATAATGCTATGCTATGGGAGCAAGTTATGGA 150
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1561 GGTTTATGACCAACTGGAT 1580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 GGATATATGATGAATTTGGAT 130
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
AZ530780/c
LOCUS      909 bp      DNA      linear      GSS 03-NOV-2000
DEFINITION Entamoeba histolytica sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION  AZ530780
VERSION     AZ530780.1 GI:11084914
KEYWORDS   Entamoeba histolytica.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica
REFERENCE  Loftus,B., Van Aken,S. and Fraser,C.
AUTHORS    Loftus,B., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica

```

```

HMI:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Class: shotgun
Seq primer: M13-Reverse
High quality sequence start: 15
High quality sequence stop: 780.
Location/Qualifiers
  1. 909
  /organism="Entamoeba histolytica"
  /strain="HMI:IMSS"
  /db_xref="taxon:5759"
  /clone_lib="Entamoeba histolytica Sheared DNA"
  /note="Vector: pHOSt; Site 1: Bst I; Constructed at The
  Institute for Genomic Research (TIGR), Rockville, MD.
  Genomic DNA isolated from broth cultures of E. histolytica
  using a method described by Clark and Diamond (Clark,
  C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
  method for isolate identification. Exp. Parasitol.
  77:450.). The DNA was mechanically sheared to give a
  tight size distribution (~2 kb). The v + i method used for
  the library construction is described in detail in Smith,
  H.O. and Venter, J.C. (Making small insert libraries for
  whole genome shotgun sequencing projects. In Genome
  Sequencing: A Practical Approach, eds. M. Vaudin and B.
  Barrell, Oxford University Press, 1999)."
  BASE COUNT      299 a 151 c 118 g 341 t
  ORIGIN

Query Match      3.3%; Score 66; DB 17; Length 909;
Best Local Similarity 48.6%; Pred. NO. 5.2e-08;
Matches 214; Conservative 0; Mismatches 220; Indels 6; Gaps 1;

Qy 1147 AACAGCTGCTGGCGCAATGACAAAGTTTGTTCAGGGAGCATACGATATCAATACCTGAA 1206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 665 AACAAATCACTCACTAATTAACAAGAGTTCTTCTCAATTAATTTGGAGAATATAAA 606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1207 GAGATTCAATATGTACAGAGAGCGCGTGATGTTGAACGGCTGGCTGATGAGGCTGCA 1266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 605 GAAATTCATTACACTGGAGCAATTAATGATCAAAATTCATGCAATTTATTACCCACCA 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1267 CAATGGAGGTGAGACAACTATCCACTTATCTTAAATACACGGCGGCTGGATATG 1326
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 545 AATATGAACAAACACAAATATCCAGTAATTTTATATACATGGAGGACCAAGATCA 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1327 ATGTACGGACATACATATTTTCAT-----GAGTTTCAGGTGCTGGCGCGAAAGGATAC 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 485 CCATGGACTAATAATTTCCATTATAGATGGAATCCACAAAGTTTATAGCACAGCAAGTTAT 426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1381 GCGTGTGTTTATATCAATCCGAGAGAGCCACGGCTAGCGGCGAGGAATTTGTAATGCG 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 425 ATTGTTTTTGGACCCCAATTTCCATGGATCAGGAAGTTATGGAGATGCAATTTTAAAGCA 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1441 GTCAGAGGAGATTTATGGGGAAGGATATGACGATGTCATGCAAGCTGTGGATGAGGCT 1500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 365 ATTAGAAGAAATTTGGGAGGATGGCCATTGGAAGATTTAATGAAGGAATGGATTTATTA 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1501 ATCAACAGAGATCCGCATATTTGATCCTTAAGCGGCTGGGTGCACGGCGGAAAGCTACGGA 1560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 305 AAAACTAGTGAGCCATTAGTTGATATAGATAATGCTATGCTATGGGAGCAAGTTATGGA 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1561 GGTTTATGACCAACTGGAT 1580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 245 GGATATATGATGAATTTGGAT 226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

142 ATTAATGCGCGCAAGGGTACTATATAGTTACGTTAAATCCAGGGGAGCAATGGATA 83
1419 CGGCGAGGAATTTGTGTAATCGGTCAGAGGAGATTTATGGGGAAAGGATTTATGAGATGT 1478
82 TAGTGAAGACTTTGCACATAAGAGTATTAAACAGAACTGGACTGGAGGACTTTCCAAGATAT 23
1479 GATGAGGCTGTGATCA 1496
22 ATTAACGGGTATTGAGGA 5

RESULT 5
BG279712
LOCUS
DEFINITION
  BG279712 Neurospora crassa sexual cDNA library, Uni-zap vector
  system Neurospora crassa cDNA clone b5e08np 5', mRNA sequence.
ACCESSION
  BG279712
VERSION
  BG279712.1 GI:13077352
KEYWORDS
  EST.
SOURCE
  Neurospora crassa.
  Neurospora crassa
  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
  Sordariales; Sordariaceae; Neurospora.
  1 (bases 1 to 475)
  Kupper, D., Lai, H., Nelson, M. and Roe, B.
  ESTs from a Neurospora crassa Sexual cDNA Library
  Unpublished (2001)
  Other ESTs: b5e08np.f1
  Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
  Department of Chemistry and Biochemistry
  Advanced Center for Genome Technology, University of Oklahoma
  620 Parrington Oval, Norman, OK 73019, USA
  Tel: 405 325 4912
  Fax: 405 325 7762
  Email: broe@ou.edu
  Contact Dr. Mary Anne Nelson, Department of Biology, University of
  New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
  regarding clone availability
  Seq primer: M13 Universal Reverse Primer
  High quality sequence stop: 474.
  Location/Qualifiers
    1..475
      /organism="Neurospora crassa"
      /strain="wild type"
      /db_xref="taxon:5141"
      /clone="b5e08np"
      /clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
      vector system"
      /tissue_type="perithecia (fruiting bodies)"
      /dev_stage="sexual"
      /lab_host="E. coli strain SOLR"
      /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
      XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
      ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT
  103 a 120 c 146 g 105 t 1 others
ORIGIN
  Query Match 2.4%; Score 46.8; DB 12; Length 475;
  Best Local Similarity 52.8%; Pred. No. 0.02;
  Matches 122; Conservative 0; Mismatches 108; Indels 1; Gaps 1;

1360 GTGCTGGCGGGAAGGATACGCGTCTTTTATATCAATCCGAGAGAGCCACGCGTAC 1419
151 GTCCTTGGGAGCAAGGTACGTCGTGTGAGCCCGCAACCCGACGAGTACGCGTTAT 210
1420 GGGCAGGAATTTGTGAA -TGGCGTCAGAGGAGATTTATGGGGAAAGGATTTATGAGATCT 1478
211 GGCATGGCTCTCCAAAACGGNAATCAAGGGCCAGTGGGTGGCGGCGGTATGAGATCT 270
1479 GATCAGCGCTGTGATGAGGTATCAACAGGAGATCCGATATTTGATCTTCAAGCGCTCGG 1538
271 GGTCAAAGCGTTTGAGCATATTGAGGAGAACATGCCGTACGTAGATACGAGCGCGGT 330

```

```

1539 TGTACGGCGGAAGCTACGGAGTTTATGACCAACTGGATCGTCGGGCA 1589
331 CGCATTTGGGTGCTAGTTATGTTGGTTATATGATCAATTTGGATCAAGGCCA 381

RESULT 6
BJ153314/c
LOCUS
DEFINITION
  BJ153314 unpublished oligo-capped cDNA library, C. elegans L1 stage
  Caenorhabditis elegans cDNA clone yk1318a03 3', mRNA sequence.
ACCESSION
  BJ153314
VERSION
  BJ153314.1 GI:18321299
KEYWORDS
  EST.
SOURCE
  Caenorhabditis elegans.
  Caenorhabditis elegans
  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
  ; Rhabditidae; Peloderinae; Caenorhabditis.
  1 (bases 1 to 697)
  Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
  and Sugano, S.
  A complementary view of the C. elegans genome
  Unpublished (2002)
  Contact: Tadashi Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
  Location/Qualifiers
    1..697
      /organism="Caenorhabditis elegans"
      /strain="N2"
      /db_xref="taxon:6239"
      /clone="yk1318a03"
      /clone_lib="unpublished oligo-capped cDNA library, C.
      elegans L1 stage"
      /sex="hermaphrodite"
      /tissue_type="whole animal"
      /dev_stage="L1"
      /note="The AD-wrncDNA library was generated with poly(A)+
      RNA isolated from both hermaphrodite and male N2 worms of
      all larval stages, embryos, adults and dauers and the
      subsequent generation of cDNAs by poly(A) priming. The
      cDNAs were cloned into pPC86"
BASE COUNT
  212 a 151 c 162 g 171 t 1 others
ORIGIN
  Query Match 2.2%; Score 43.6; DB 13; Length 697;
  Best Local Similarity 51.3%; Pred. No. 0.23;
  Matches 100; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

1375 GGATACGGTCGCTTATATCAATCCGAGAGAGCCACGCTACGGGCGAGAAATTTGTG 1434
696 GGATATGCCCTNTCAAGTCAATTCGCTGGATCACTCGGATTTGGTACGATTTATT 637
1435 AATGCGGTCCAGAGGAGATTATGGGGAAAGGATTATGATGATGATGATGATGATGAT 1494
636 CGTGCTCGCTGGAAACTCGGAGATATGGATGTTAAGGATGTTTATAACGCGATCTC 577
1495 GAGGCTATCAACAGAGATCCGCATATTTGATCTTAAGCGGTCGGTGCACGGCGGAAGC 1554
576 ACCGTCTTGGATGAAGATCCACGTTATTTCTCGAGACAAGGTTGTTACTCTTCGGAGGATCT 517
1555 TACGGAGGTTTATG 1569
516 CATGGAGGGTTCCTG 502

RESULT 7
BJ118680/c
LOCUS

```

DEFINITION

EST066 Differentially expressed cDNA libraries of BPH-treated/blast fungus-infected rice leaf tissues *Oryza sativa* cDNA clone BfHN-n31, mRNA sequence.

ACCESSION

Bf118680

VERSION

Bf118680.1 GI:14571312

KEYWORDS

EST.

SOURCE

Oryza sativa.

ORGANISM

Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; *Oryza*.

REFERENCE

1 (bases 1 to 840)

AUTHORS

Song, F.M. and Goodman, R.M.

TITLE

Molecular cloning of differentially expressed cDNAs associated with systemic acquired resistance in rice

JOURNAL

Unpublished (2001)

COMMENT

Contact: Song, F.M.; Goodman, R.M.
Department of Plant Pathology
University of Wisconsin-Madison
Rm 689, Russell Laboratories, 1630 Linden Drive, Madison, WI 53706, USA
Tel: 608 262 9162
Fax: 608 262 8643
Email: fmsong@plantpath.wisc.edu; rgoodman@facstaff.wisc.edu
Benzothiadiazole- and/or Magnaporthe grisea-inducible
Insert Length: 840 Std Error: 0.00
Seq primer: M13 Forward and reverse.

FEATURES

source

1..840

/organism="Oryza sativa"
/cultivar="Yuanfengzao (a susceptible variety to rice blast disease)"
/db_xref="taxon:4530"
/clone="BfHN-n31"
/clone_lib="Differentially expressed cDNA libraries of BPH-treated/blast fungus-infected rice leaf tissues"
/tissue_type="Seedling leaves"
/dev_stage="Three-week-old rice seedlings pretreated with BPH and/or inoculated with Magnaporthe grisea"
/note="Vector: pGEM-T Easy; Rice seedlings were pretreated by foliar spraying with 0.3 mM benzothiadiazole (BTH) and then inoculated with Magnaporthe grisea three days after BPH treatment. Leaf samples were collected twenty-four hour after inoculation and mRNAs were used for construction of the differentially expressed cDNA libraries by suppression subtractive hybridization."

BASE COUNT

ORIGIN

Query Match 2.2%; Score 43; DB 13; Length 840;

Best Local Similarity 49.3%; Pred. No. 0.39;

Matches 112; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 1284 AACATATCCACTTATCTTAACATACACGGGGTCCGATATGATGTCAGGACATACATA 1343

Db 840 AACAGACACAAATGTGTCTTCATGGTGGACACACACTGTTCACCATCAAGCTA 781

QY 1344 TTTTCATGAGTTTCAGGTGCTGGCGGCGAAGGATACGCGGTCTTTATATCAATCCGAG 1403

Db 780 TTCAAATCGTTAGCATTTCTTATTCACGGGATATAACCTCTGTGTGAACTATAG 721

QY 1404 AGGAAGCCACGGCTACGGCGAGGAATTTGTGAATCGGTACAGAGAGATTATGGGGGAAA 1463

Db 720 AGGTTCTACTAGCTTTGGTGAAGAACACTACAACTCTCTCTCGGAATATTTGTTCTCA 661

QY 1464 GGATTATCAGCATGTGATGACGGCTGTGGATGAGCGCTATCAACGAG 1510

Db 660 GGATGTGAATGATGTATTGACGGCTTTGGACTTTTGTATATAAGAAAG 614

RESULT 8

AZ528669/c

LOCUS

AZ528669 873 bp DNA linear GSS 03-NOV-2000

DEFINITION

ENTC042TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION

AZ528669

VERSION

AZ528669.1 GI:11081093

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica.

ORGANISM

Entamoeba histolytica.

REFERENCE

1 (bases 1 to 873)

AUTHORS

Loftus, B., Van Aken, S. and Fraser, C.

TITLE

Determination of clone end sequences from Entamoeba histolytica

JOURNAL

HML:IMSS sheared DNA library

COMMENT

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjlloftus@tigr.org

Clones are derived from the Entamoeba histolytica HML:IMSS sheared DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 17

High quality sequence stop: 849.

FEATURES

source

1..873

/location/Qualifiers

/organism="Entamoeba histolytica"

/strain="HML:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 Kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 327 a 132 c 101 g 313 t

ORIGIN

Query Match 2.2%; Score 42.6; DB 17; Length 873;

Best Local Similarity 54.0%; Pred. No. 0.53; 74; Indels 0; Gaps 0;

Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1420 GGGCAGGAATTTGTGAATCCGTCAGAGAGATTATGGGGAAGGATTATGACGATGTG 1479

Db 872 GGAGATGCATTTTAAAAAGCAATTAGAAAGATTGGGAGGATGGCCATTTGAAGATTTA 813

QY 1480 ATGCAGGCTGTGGATGAGGCTATCAACAGAGATCCCATATGATCCTTAAGCGGCTCGGT 1539

Db 812 ATGAAGGAATGGATTTATTTAAACTTAGTGAGCCATTAGTTGATATAGATATGATGATGT 753

QY 1540 GTCACGGCGGGAAGCTACGGAGGTTTATGACCACTGGAT 1580

Db 752 GCTATGGGAGCAAGTTATGGAGGATATATGATGAATGGAT 712

RESULT 9

BG025910

LOCUS

DEFINITION

mRNA sequence.

ACCESSION

BG025910

VERSION

KEYWORDS

BG025910 939 bp mRNA linear EST 24-JAN-2001
602292057F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4386711 5',
mRNA sequence.
BG025910.1 GI:12412979
EST.

